

US politics: campaigns as usual

Neither the Republican nor Democratic party has produced a 'platform' spelling out a policy agenda for science and technology in any meaningful detail.

THE end of August brings with it the blessed end to the quadrennial summer of political conventions and writing of party 'platforms' as the Republican and Democratic candidates for the White House try to define themselves and distinguish themselves from their rivals. The Democratic platform, a clear statement of the goals of candidate Bill Clinton of Arkansas, gets right down to business with talk about investment in science and technology, education and the deficit. The Republican treatise, following President George Bush's lead, casts the party as the bastion of "family values". "As the family goes, so goes the nation."

It is well known that political platforms are not gospel, that they deal as much in rhetoric as in fact and detail. Nevertheless, they provide clues to the parties that set the tone for the campaign and, this year, reveal some striking changes in thinking, especially among the Democrats.

Taking a cue from former candidate Paul Tsongas of Massachusetts, who entered the campaign telling the Democrats that you cannot redistribute wealth that has not been created, the final version of the Democratic platform "rejects" the "big government theory that says we can hamstring business and tax and spend our way to prosperity". Sounding very Republican, the Democrats attest to a belief in "free enterprise and the power of market forces," aided by a "national economic strategy". Whether that means that Clinton really favours a formal economic or industrial policy remains to be seen.

But some points are spelled out. Clinton wants Congress to agree to reinvest defence savings in research and education (assuming that the so-called 'peace dividend' is real), and proposes an investment tax credit and a reduction in capital gains taxes (sounds like Bush) for "patient investors in emerging technologies and new business". But, in keeping with the Democrats' image as the party of the common man, Clinton has also proposed an unspecified tax increase for the wealthy, defined as an income of more than \$200,000 a year. The Democrats also think innovation and the US's competitive position *vis-à-vis* Japan and Germany deserve prominent mention. The party's second plank calls for a permanent research and development tax credit, and a doubling of basic research budgets for unidentified "key technologies".

By contrast, the Republican platform devotes its first several pages to the family, calling on good Republicans to "oppose and resist" Democratic party efforts to redefine the traditional family. That is to say, the Republicans reject the idea that homosexuals can form a family, and that they deserve

special protection under civil rights laws. The Republicans officially deplore abortion and fetal tissue transplantation research, and propose reestablishing the "maternity homes" where unwed mothers used to go before the birth of a baby who was usually given up for adoption. The Republican platform rejects the distribution of condoms and clean needles as part of a programme to reduce the spread of AIDS, instead stressing abstinence and marital fidelity.

Although the emphasis and positioning in the platform is different, the Republicans express a belief in the connection between science and technology and increased economic productivity, with a list that covers the waterfront. Telecommunications, high-performance computers, high-speed data networks, digitization, advanced software, energy, agriculture, oceanography and geological research are among dozens of items that get a nod, though how increased research would be paid for is not mentioned.

Indeed, in his acceptance speech, Bush called for an economy-stimulating across-the-board tax cut (presumably income tax) coupled with reductions in federal spending in all areas except Social Security. However, he refused to give a single specific example of what he has in mind.

Taken together, the platform generalities and the candidates' rhetoric tell us little that is really substantive. Both Bush and Clinton want to end the economic downturn in the United States. Each wants strong families, though neither fully knows how to create them. (Here, it might be said, Clinton's idea of creating a government-funded work programme to create jobs rebuilding roads, bridges, and parks — reminiscent of Franklin D. Roosevelt — does offer one specific approach to employing youth.) After two months of talk from each side, instead of coherent, fully formed policy proposals, the United States is facing a fall campaign with an emphasis on name-calling and casting blame — Bush on the Democratic Congress that has been in power for the past 12 years, Clinton on the Reagan and Bush White House.

Meanwhile, serious challenges abound. For instance, the value of the dollar reached a new low against the German mark last week (the lowest since the Second World War) in response to US interest rates that the Federal Reserve Bank has pushed as low as 3 per cent to stimulate the economy while European nations, particularly Germany, are fighting inflation with interest rates as high as nine per cent.

Is it too much to hope for a presidential campaign focused on the national and international issues that are of substance? It would seem so.



Wang goes bust

The bankruptcy of Wang Laboratories is sad — and confirmation that even the best cannot stand still.

A DECADE ago, none of the glossy business magazines failed to publish a profile of the late An Wang, who, despite his birth in China, had founded Wang Laboratories in the United States and, in the face of indigenous competition, had turned it into a substantial company (the sales, at the peak, exceeding \$3 billion a year). Wang, the man, had become a legend in his time. At the outset his stock in trade, before the microprocessor revolutionized the computer business in the late 1970s, was to offer commercial businesses access to computer techniques of a kind more commonly found in research laboratories. For a time, Wang installations in office buildings were recognizable by their sophistication. So why, last week, did this paragon of technology sue for bankruptcy under the rules of Chapter 11 of the US bankruptcy legislation (which get creditors off a company's back while it tries to reorganize itself)?

The simple answer is that Wang has been losing money — \$385 million last year, \$130 million in the most recent financial year (to the end of June). Like many other technologically advanced companies, Wang sustained its growth over two decades by borrowing from the financial markets; the bond holders not unreasonably want to be paid the interest they are owed although the company is losing money. The explanation of Wang's plight can be sought at several different levels. Technically, it is a fact that the minicomputers that were the hard core of its hardware business have been outflanked by desktop machines. (Digital Equipment Corporation, as recently as a decade ago applauded for its VAX minicomputers, is now also in trouble.) The management consultants were also shaking their heads over the degree of An Wang's personal identification (until his death last year) with the company's past success. But who believes that committees of directors are able to respond more quickly than individuals to changing circumstances? The truth is that Wang's management of the company is blamed now simply because it has failed.

What will happen next is anybody's guess. The company hopes to find a continuing role as a superior designer of integrated software systems, but that must be a gamble. For one thing there is already a lot of competition. For another, it is improbable that major companies outside the computer business will indefinitely delegate to outsiders, however skilled, the design of the information management systems on which their own effectiveness depends. But the court supervision required by Chapter 11 would probably not allow Wang to use its technical excellence to strike out in a more adventurous direction.

The wider lessons of Wang's collapse are more deeply buried. Of course, in retrospect, it was a mistake to have made computer systems that were incompatible with those of other manufacturers, but suppose that gamble had paid off? Wang's more serious mistake was to have to set out, in

a field in which technology was plainly destined to be quickly changing, to become an integrated company embracing research, development, manufacturing and the selling of the final products. Those are the circumstances in which companies risk being trapped by newly created inertia into products rendered obsolete for reasons beyond their own control. In the United States, the biotechnology industry (or those who invest in it) seems to have sensed the danger, and has remained a collection of research-based companies selling licences (and sometimes themselves) to bigger fish. Much of what is called Silicon Valley seems to have taken the same point. Those who worry about the industrial competitiveness of the United States may find it some comfort that the importance of research-based companies is almost unique to the United States. □

Darwin's detractors

Those persuaded that natural selection is a pack of lies deserve no help from the serious press.

DARWIN's theory of evolution has never sat comfortably in the minds of those who pretend that science is a myth engendered by the malevolent to disturb the innocent among us. So it is natural, and inevitable, that there should be a steady stream of publications intended to demonstrate to those who already believe evolution to be false that their prejudices are correct. The latest version of the attack on Darwin, *The Facts of Life* (Fourth Estate, London) is written by one Richard Milton, who is not a creationist. Like others in the field, the book, as the saying goes, seems to be "well researched"; it quotes the experiments of Cairns *et al.* (*Nature* 335, 142–145; 1988) as proof of directed rather than undirected evolution, for example.

The appearance of books like this is not remarkable; publishers are free to publish what they choose. But Milton's book has this week been given advance publicity in the British newspaper *The Sunday Times* under the headline "Scientist threatens to make Darwin extinct". (Milton is described in the text that follows as a "science writer, engineer and amateur geologist"; this is his first book.) The text of the article, taking more than a third of a page, summarizes the "evidence" that the Earth is much younger than commonly supposed. In the tradition of balanced journalism, it also includes supportive comment from a retired bishop and expressions of mild dissent from two geneticists.

Why serious newspapers do this kind of thing is beyond belief. To be sure, these are the dog days in Britain, when domestic politics is in limbo and newspapers preoccupied with the remarkable doings of the royal family. No doubt the newspaper, if asked, would explain that it is merely its public duty to bring to the attention of its readers controversies that, if resolved, would change the intellectual temper of society. But, followed uncritically, as in this case, that is simply a licence to revive spurious and outdated arguments and to dress them up as alive whenever the flow of regular news is slack. That is not a service but a disservice to serious readers. □

Rising toll from AIDS in Japan prompts big jump in spending

Tokyo. Japan is at last waking up to the dangers of AIDS. The Ministry of Health and Welfare last week responded to a sharp rise this year in the number of reported cases of infection by human immunodeficiency virus (HIV) by requesting a fivefold increase in fiscal year 1993, starting on 1 April, for its tiny budget to fight AIDS. Nearly half of the proposed budget would go to research, with the rest invested in public education, screening, treatment and counselling.

But critics say that even a fivefold increase falls far short of what is required. This year, the government will spend only a little more than ¥2 billion (US\$16 million) to combat AIDS.

Japan has few AIDS patients compared with the United States, Europe and some South-East Asian nations. The official total, including those who have died, stands at just under 500, but many cases are said to go unreported because of fear of discrimination. However, in the first six months of this year, 226 new cases of HIV infection were reported, almost the same number as in the whole of last year.

A change in the affected population is also driving the increased support. Many of those newly infected are believed to have contracted the virus through heterosexual sex, whereas in the past the majority of reported cases were among haemophiliacs and homosexuals. This development has heightened public concern and awareness of AIDS.

The government made its first major investment in fiscal year 1988 after the first Japanese woman died of AIDS. But public concern subsided rapidly and the annual budget has remained frozen at ¥2 billion without even increases to cover inflation.

Next year, however, the ministry would like ¥10.3 billion. The request has to be screened by the Ministry of Finance and approved by the Diet, but such requests seldom undergo significant change.

Nearly half of the budget, ¥4.8 billion, would go to research and the development of drugs and vaccines; this figure includes ¥576 million for the World Health Organisation. One of the nation's major centres for AIDS research is at the National Institute of Health, whose researchers are moving into a controversial new building (see page 703). Shudo Yamazaki, director of the centre, says that his institute asked for a tenfold increase but expects to receive less than half that amount.

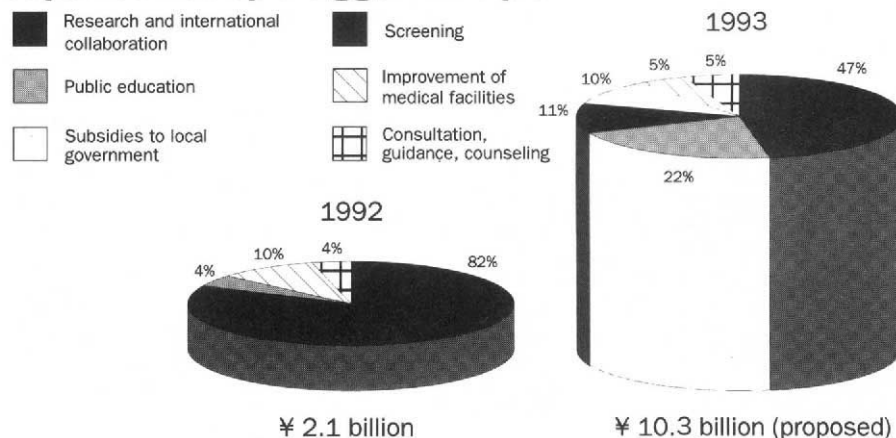
A significant portion of the increased budget will be used for public education. Ignorance and prejudice about AIDS are widespread: many hospitals refuse to treat those with AIDS

for fear of frightening away other patients, while others refuse to acknowledge that they care for patients with HIV.

Only two hospitals, both in Tokyo, spe-

The health ministry provides the lion's share of government funding for AIDS research, although the Ministry of Education, Science and Culture (MESC) and the Sci-

Japan carves up a bigger AIDS pie



cialize in the treatment of AIDS patients — the Institute of Medical Science of Tokyo University and Tokyo Komagome Hospital. The ministry would like to establish similar centres in other parts of the country. Officials also plan to establish Japan's first hospice for the treatment of terminally ill AIDS patients, although critics say that the 20-bed facility falls woefully short of what is needed.

ence and Technology Agency (STA) also provide some support. Although these two agencies are said to be considering ways to increase their support, any increases are likely to come at the expense of other programmes, such as the human genome project. Their intentions will be spelled out in 1993 budget requests due in the next few weeks.

David Swinbanks

Ways to keep Mars 'clean'

A report from the US National Research Council has allayed fears that a proposed European mission to Mars might be jeopardized by demands for stringent standards of planetary protection (see *Nature* 358, 443; 1992). The report*, commissioned by the US National Aeronautics and Space Administration (NASA) in advance of next week's World Space Congress in Washington, says that probes landing on Mars need not be sterilized as long as they are not conducting exobiological experiments.

Researchers at the European Space Agency (ESA) feared that such a requirement would break the budget for its proposed Marsnet project, which is intended to study the planet's meteorology and seismic activity. Because its mission is non-biological, however, the project would require only a thorough cleaning during assembly.

Ken Neilson of the University of Wisconsin-Milwaukee, chair of the committee that produced the report, says that the committee feels that any contamination from a

non-exobiological probe such as Marsnet is unlikely to spread beyond the landing site. If such a probe discovered an environment favourable to life, a subsequent probe could examine that possibility by landing beyond the possible sphere of contamination.

However, the planetary protection standards are likely to get tougher for probes whose main purpose is exobiological. The report says that the increased sensitivity of methods to detect biological material developed since the Viking missions of the 1970s demand equally stringent sterilization procedures.

The report also emphasizes the importance of gathering data from a range of martian sites. "Until such data are available", it says, "it will be impossible to make informed decisions concerning landings for in-depth biological study." **Ian Mundell**

* *Biological Contamination of Mars — Issues and Recommendations* (Space Studies Board, National Research Council, 2101 Constitution Avenue NW, Washington, DC 20418).

NSF's proposed conflict-of-interest rules place burden on applicant and institution

Washington. US researchers would be required to disclose all relevant financial ties when submitting a grant application as part of a conflict-of-interest policy under consideration by the National Science Foundation (NSF). The proposed regulations, published on 16 July in the *Federal Register*, would also require institutions to have a comprehensive policy to deal with such potential conflicts.

If approved, the regulations would constitute the first such conflict-of-interest policy to be adopted by a US science agency. The National Institutes of Health withdrew a version they had circulated in 1989 after the academic community protested that it would cost millions of dollars to implement and would duplicate institutional policies in place at most universities (see *Nature* 343, 104; 1990).

NSF, however, may have an easier time of it. Over the past three years, Congress has repeatedly urged the science agencies (and the academic community) to address the issue, despite the feeling that instances of significant conflicts of interest are rare. Most researchers "can see the writing on the wall", says one NSF official, and have accepted the idea that regulations are inevitable.

If such regulations must come, the proposal from the NSF may be the most palatable. As written, it is a way to gather data rather than an attempt to police research. According to the draft regulations, "NSF believes that significant conflicts of interests are rare in the work that NSF funds, [but] we have no data by which to demonstrate the point. The proposed policy would create a bank of data to serve as a reality check."

NSF expects disclosure (and a review by the institution) in cases where proposed research would evaluate or lead to a commercial product or where the research would be relevant to any outside commercial venture in which the investigator is involved. But it believes that few researchers will fall into either category, and that those who do should find it simple enough to disclose that fact. It estimates that the typical investigator should need only 20 minutes to complete the financial information.

The agency intends that the internal policies of universities and other institutions that receive NSF funds should deal with conflicts. It has not yet decided how it will handle a case of conflict that the applicant's institution has not adequately dealt with, or, for that matter, how it will identify such a case.

NSF would require institutions that receive grants to have a conflict-of-interest policy that, at the minimum, requires

faculty and staff to disclose any 'significant' financial ties. Such a policy would include procedures to review any disclosures and to inform research sponsors both of any problems and how they have been resolved.

Under the proposed regulations, investigators would submit financial disclosures in a sealed envelope accompanying their grant application. The application would be reviewed for its scientific merit and the envelope opened only if the application were approved for funding. At that point, NSF's general counsel would review the disclosure for any possible conflicts using criteria yet to be drawn up.

The proposal has so far elicited little reaction. Part of that silence is due to its appearance at a time when many institutions are on holiday. The original deadline of 14 September has been extended to 14 October at the request of the Association of American Universities (AAU), an associa-

tion of 58 major researcher universities in the United States and Canada, which plans to discuss the issue at a meeting on 8 September.

Those who have read the proposed regulations see few problems. Milton Goldberg, of the Washington-based Council on Government Relations, points out the ambiguity in the phrase "significant financial ties" and is concerned that the disclosures may become public through requests made under the federal Freedom of Information Act. NSF says it would try to block such requests.

But even Goldberg says that it is important to "separate the nuisance [to researchers] of doing this and its public-policy value". Getting hard data on the scope of the problem may clarify the issue, he says, and is worth some potential loss of privacy and inconvenience to scientists and their institutions.

Christopher Anderson

Australia increases funds for grants and applied research

Sydney. The Australian government will increase research spending next year by 2.2 per cent, according to budget figures released last week. The biggest winners are university scientists funded by the Australian Research Council, which will receive an increase of 9 per cent, while most government researchers will see their funding actually shrink. The overall science budget will grow to A\$2.85 billion (US\$2.1 billion), an increase of A\$149 million over 1992.

The government's budget is rarely changed once it is presented. And as the government is by far the largest supporter of science, its priorities tend to shape the direction of research in Australia.

For industry, the best news is the government's decision to abandon its plans to weaken the incentive for companies to invest in research. Cabinet ministers had decided to reduce the 150 per cent tax deduction that companies can claim for money spent on research and development, but the decision was reversed just days before the budget was unveiled after strong protests from the business and research communities.

In place of a smaller deduction, the government will tighten its rules to restrict the tax concessions to genuine investors in research projects, excluding those looking solely for a tax shelter. The government estimates that it will lose A\$325 million

because of the deduction, which is intended to stimulate investments in research that will eventually bolster the economy.

The budget for the research council, which dispenses most of the federally funded grants under the country's centralized funding system, will increase to A\$269 million. Funds for agricultural research will rise by 13 per cent, to A\$110 million. The budget for cooperative research centres has been tripled, to A\$54.5 million, to accommodate several new university industry centres.

In contrast, the government continues to pare down its own research operations. The largest component, the Commonwealth Scientific and Industrial Research Organization (CSIRO), will have its budget reduced by 1 per cent, to A\$464 million. The CSIRO is expected to earn another A\$220 million from other sources; as a sign of its increasing reliance on the business community, it is moving its headquarters from the isolated federal capital of Canberra to Melbourne, the country's second-largest business centre.

The government has also issued a white paper (policy document) stressing the importance of applied research. The Prime Minister's Science Council is being renamed the Science and Engineering Council, and A\$6 million will be spent to create three advanced engineering centres to bring together university and industrial researchers.

Mark Lawson

NASA ponders "faster, cheaper" way to send scientific mission to Pluto

Washington. The new head of the US National Aeronautics and Space Administration (NASA) has thrown his weight behind a proposed mission to Pluto that would demonstrate the feasibility of his philosophy of "better, faster, cheaper". The mission, which would be the first to reach the most distant planet in the Solar System, is moving rapidly through the ranks of the agency and may even appear this winter as a new start in the agency's proposed budget for the fiscal year 1994 that begins on 1 October, 1993.

Mission designers at the Jet Propulsion Laboratory (JPL) in Pasadena, California, have developed a concept for a 150-kg spacecraft that could travel the nearly 5 billion km to Pluto in as few as seven or eight years. The mission, called the Pluto Fast Fly-by, is small compared to the 800-kg Voyager probes or the 1,500-kg Galileo now on its way to Jupiter and relies on brute force to accomplish its goal. It would use the largest expendable launcher in the US fleet, the Titan 4/Centaur, to send two nearly identical spacecraft on a direct path to Pluto.

The onboard science instruments would be kept to a minimum: most likely a camera, infrared and ultraviolet spectrometers and a radio science experiment. The second spacecraft would arrive months after the first fly-by, providing a backup and allowing both hemispheres of the planet and its satellite Charon to be photographed at close range. Each spacecraft would observe Pluto and Charon for about a month, with the closest encounter lasting only a matter of hours.

What distinguishes this mission from other ideas waiting for funding is the strong interest shown in it by Daniel Goldin, who succeeded Richard Truly in April as administrator of NASA. Goldin was so taken with the idea after hearing it described during a visit to JPL in May that he directed NASA's Solar System Exploration Division to speed up its studies. What had been a typical review was turned into a 60-day study.

That timetable may allow it to be included in the 1994 budget request that NASA will submit next week to the White House. It would be the first in a new class of intermediate-sized missions, says Wes Huntress, head of NASA's planetary division.

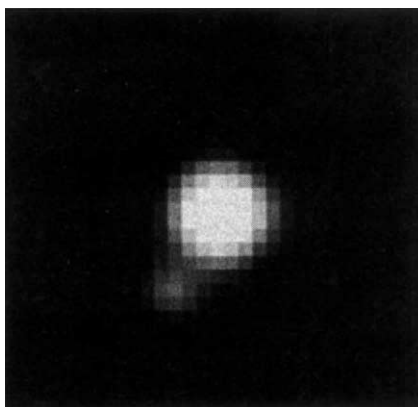
That approach is seen as a way to cope with Washington's current climate of fiscal austerity, a situation that earlier this year caused the cancellation of an ambitious mission known as Comet Rendezvous/Asteroid Fly-by (CRAF). The agency has proposed a new 'Discovery' series of low-budget missions, including a small Mars lander and a

rendezvous with a near-Earth asteroid, that would cost no more than \$150 million each.

The cost of intermediate-class missions would be a maximum of \$400 million, says Huntress, not counting the launch vehicle, which in the case of a Titan/Centaur adds another \$350 to \$400 million. (The agency

some time around 2015 as the planet recedes from the Sun. In addition, some 40 per cent of Pluto's surface will be in permanent shadow by 2020, a function of its pole being pointed towards the Sun.

A group headed by planetary scientist Alan Stern at the Southwest Research Insti-



The best views of Pluto and Charon, its moon, from Earth, left, and from the Hubble Telescope, right, are still a far cry from what scientists want to see.

could save money by using a Soviet Proton launcher, if it can resolve certain technical issues.) By comparison, the budget for the proposed Cassini mission to place an orbiter around Saturn rose to \$1.7 billion before NASA decided to scale back the project last spring.

One way to hold down the cost of the Pluto mission would be to take advantage of recent advances in the miniaturization of spacecraft components, including the use of a star tracker developed for the Pentagon's Strategic Defense Initiative that weighs one-tenth as much as similar devices used on past interplanetary spacecraft. Making this and other lightweight hardware durable enough for a ten-year flight into deep space would significantly reduce the mission's cost. The project also plans to borrow a leftover communication antenna from the Viking project to Mars in the 1970s.

Perhaps the most significant savings would come from speeding up the development cycle for the project and keeping the engineering teams small. If work begins as soon as 1994 or 1995, the spacecraft could be ready for launch as early as February 1998 — less than half the development time for such recent missions as the Hubble Space Telescope and Galileo.

There are good scientific reasons to launch the Pluto fly-by mission as soon as possible. Researchers would like to study Pluto's tenuous atmosphere before it disappears

tute in San Antonio, Texas, has studied several proposed missions to Pluto during the past several years, but none has been this far along in the agency's approval process. Two years ago, the group for a time considered a large, Cassini-sized spacecraft before it was judged to be too big, slow and expensive. Alternatively, a proposal for a tiny, 35-kg Pluto spacecraft equipped with a camera and little else was dismissed as not being big enough. The fly-by strikes a happy medium, says Stern. "We don't want to do a 'photo op'", he says. "We want to do first-rate science, and the 150-kg mission satisfies our highest scientific priorities."

Stern believes that missions such as the Pluto Fast Fly-by are the best hope for an active US planetary programme. "We were asked to do things differently and we did", he says. "I think this mission will be a test of whether the US wants to remain in the planetary business."

Tony Reichhardt

Correction

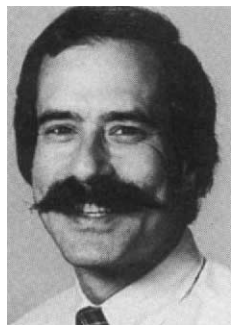
A CONTAMINANT in a batch of L-tryptophan being investigated as the cause of an outbreak of eosinophilia-myalgia syndrome (EMS) was incorrectly described in a recent article in *Nature* (358, 96; 1992). Its correct name is 3-phenylamino-L-alanine. Researchers at the National Institute of Hygienic Sciences in Tokyo first discovered the contaminant in high-performance liquid chromatographs.

Court-appointed scientists provide technical expertise

Washington. Judges at all levels of the US legal system have begun using court-appointed researchers to help them resolve contentious technical issues. Such impartial third-party experts serve as a counterweight to 'expert witnesses' hired by either plaintiffs or defendants by offering impartial advice to clarify issues for judges and juries.

A recent case demonstrates the power of a court-appointed expert. In 1989, Computer Associates International Inc. (CAI) claimed that Altai Inc. had copied a CAI computer program, thereby violating software copyright laws.

To untangle the technical details, US Circuit Court Judge George Pratt of New York called in Randall Davis, associate director of the artificial intelligence laboratory at the Massachusetts Institute of Technology.



Dr Randall Davis

Davis studied the legal criteria used to evaluate a computer program's originality — and found that those criteria, set by a judge in 1976, were "incoherent". Pratt agreed and proposed a new standard for appraising possible copyright violations. In June, a New York appeals court upheld Pratt's decision.

Such court-appointed experts are neutral informants paid to help juries and judges understand the scientific details of product liability lawsuits, patent cases and other highly technical cases. A series of asbestos lawsuits in the late 1980s illustrates what court-appointed experts do and their considerable credibility.

Confused by conflicting testimony, a judge for the US district court in Cincinnati asked neutral pulmonologists to testify about whether the plaintiffs suffered from asbestos-related disease. Although juries tend to side more with the plaintiffs in such cases, the juries in these cases followed the opinion of the court experts in 13 of the 16 cases and awarded money to only four plaintiffs.

The appearance of such court experts is still the exception rather than the rule. A study by the Federal Judicial Center (FJC), a government research institute serving the judiciary, found that only 20 per cent of federal district court judges have ever relied on a court-appointed expert. Moreover, half of those judges had appointed an expert only once.

The most frequently appointed experts

were medical professionals, followed by computer engineers and program designers. Fewer than 10 per cent were natural scientists, according to the FJC study.

Although few judges rely on court-appointed experts, many have occasionally wished for help in sorting out conflicting testimony. Judges are particularly baffled by cases involving DNA testing, by toxic torts—suits claiming that a substance causes cancer or other health problems—and by lawsuits over biotechnology and computer patents. When the plaintiff's experts and the defence's experts completely contradict each other on arcane technical issues, judges and juries do not know whom to believe.

Court-appointed experts perform a variety of duties to help break down such legal logjams. Sometimes court experts act only as advisers, defining the terms and basic principles of statistics or neurology. In other cases, they comment on the evidence presented by the two sides.

Few judges allow their court-appointed experts to tackle issues as fundamental as those examined by Davis. Part of the reason is their belief that court-appointed experts strike a blow at the adversarial system, one of the foundations of US jurisprudence. A 'neutral' third witness, they fear, might disrupt the plaintiff-versus-defence scheme for organizing trials and make it possible for juries to disregard the evidence presented. Court-appointed witnesses also make it harder for each side to control what evidence is offered.

But some judges have another, more practical, reason for being wary of court-appointed experts: science does not always provide answers that the courts can use. If a judge wants to know whether a chemical causes cancer, a call for more research will not suffice.

"I don't have a theoretical decision" to make, says Judge Carl Rubin, who presided over the asbestos suits. "I have a lawsuit. Is it more than a 50 per cent chance that A causes B or not? That's all we can ask."

In addition, some cases simply cannot be boiled down to scientific questions. "If people living next to a toxic dump site are claiming all kinds of injuries, to a scientist it looks like a scientific issue," says Margaret Berger, a professor of law and science at Brooklyn Law School. "To a judge it looks much more like a social problem."

Nonetheless, many scientists and judges agree that a wise use of court-appointed experts can contribute to solving difficult cases. And as cases resting upon high-technology grow in number and complexity, it is likely that more judges will rely on court-appointed experts.

Traci Watson

CHINA BRIEFS

Beijing. China will soon begin a nationwide survey of its wetlands in the wake of its acceptance of an international convention to preserve these important habitats. China's wetlands, the largest in Asia, are being eyed hungrily by developers, and its decision to become the 67th country to ratify the global convention is seen as a way of marshalling outside support for its conservation efforts. The survey by the forestry ministry, which is responsible for about 90 per cent of the country's wetlands, is expected to provide a forum for cooperation between several groups interested in preserving biodiversity that are not now working together. But there are no provisions in the survey for new research projects.

Beijing. China signalled its willingness to abide by international rules protecting intellectual property last month by joining the Universal Copyright Convention. China has had a poor record of respecting the ownership of intellectual property, and pirated versions of foreign-language books and products are commonly sold in shops marked with signs saying "no admission to foreigners". Its endorsement of the global convention means that from October it will pay for foreign copyrights with foreign currency. China's National Copyright Administration is drafting regulations to carry out the convention, and the agency that handles international traffic has expanded its services to include Britain, the United States and countries of the former Soviet Union.

Beijing. A record number of archaeologists from around the world have applied to China's State Administrative Bureau of Cultural Relics this year for permission to study Chinese artefacts. This surge of interest follows recent reports of the discoveries of human crania found in Middle Pleistocene terrace deposits of the Han River and a decorated antler fragment from a 13,000-year-old Upper Palaeolithic occupation layer in Hubei Province (*Nature* **357**, 404 & **356**, 116; 1992). This spring, a French team was granted permission to work in the Xinjiang Uyghur Autonomous Region in northwestern China and a Canadian team was allowed to work in Heilongjiang Province in northeastern China. The cooperative projects are mostly those on which Chinese scientists have already done research, according to Wang Li-Mei, an official with the cultural relics bureau, and the excavations have focused on such ancient sites as the Xinjiang desert. The projects provide China with foreign funds and give its researchers access to advanced techniques. One project is a \$4 million relics protection and restoration centre in Xian, in Shaanxi Province, to be funded by the Italian government. The region is surrounded by such well-known archaeological sites as the terracotta tombs and the great tomb of China's first emperor.

You Qin Li

Japanese open infectious disease lab despite local and national protests

Tokyo. This week, after waiting 20 years, hundreds of scientists at Japan's National Institute of Health will begin moving from their dilapidated headquarters in Tokyo to a plush new building in another part of the city. The move has angered local residents, who fear that an accidental leak of contagious agents would harm a densely populated region that contains a hospital. And the construction itself has unearthed cultural artefacts as well as possible evidence of atrocities committed by Japan during the Second World War.

The institute, with more than 300 researchers, is the nation's leading centre for research into and control of infectious diseases. It is in a shabby, 60-year-old building that once belonged to the now defunct Imperial Navy. The walls are cracking, the building looks as though it has never been given a fresh coat of paint and the dark corridors are grim even by the standards of Japan's notoriously drab national laboratories. Yet the institute plays a vital role in the control of infectious disease, much like the US Centers for Disease Control in Atlanta. The institute also screens the safety of vaccines, antibiotics and other biological products as well as carrying out basic research into infectious diseases.

The Ministry of Health and Welfare, which administers the institute, initially favoured a site in Tsukuba science city north of Tokyo. Although the plan was approved in the early 1970s by the Diet, researchers fought hard against being sent to what was then just a collection of national laboratories in the middle of the paddy fields of Ibaraki Prefecture.

The Tsukuba plan was abandoned, and in 1986 the ministry announced a new plan to move to a site next to the National Medical Centre in Shinjuku Ward formerly occupied by a centre for handicapped children. But this idea has also encountered opposition.

The new site is in a densely populated area and local people are worried that a laboratory conducting experiments on the AIDS virus and other contagious agents could leak into the local environment. The laboratory will be at Level 3 on the internationally used biosafety scale of 1 to 4, where 4 indicates a facility for handling large amounts of the most contagious agents.

According to Shudo Yamazaki, deputy director-general of the institute, the Level 3 laboratory will have double the necessary number of high-performance air filters capable of removing contagious agents as small as 0.3 micrometres. The new building has also been designed to stand up to an

earthquake comparable to the one in 1923 that flattened much of Tokyo.

Residents are demanding assurances that the building is completely safe, says Yamazaki, and opposition continues despite dozens of meetings with local residents. In 1989, residents filed suit in the Tokyo district court with the help of Shingo Shibata, a professor of social science at Hiroshima University. But the case is expected to languish for years in the grossly undermanned Japanese court system.

Local opposition is not the only problem facing the new facility. In 1987, shortly after construction began, pots and pans from the Yayoi Era (about 300 BC to AD 300) were unearthed and construction was halted for several months to allow excavation. When construction resumed, riot police were used to quell a protest by a small group of students from nearby Waseda University, an action that still troubles residents and university officials.

As if that were not enough, in 1989 construction workers unearthed dozens of recently buried human bones of unknown origin. There is speculation that the bones

are the remains of Chinese victims of biological experiments carried out in Manchuria during the Second World War — the site is the former location of the headquarters of a medical corps of the Japanese Imperial Army that performed the experiments.

The Ministry of Health and Welfare is

The old (right) and new (below) homes of the National Institute of Health in Tokyo.



studying the origin of the bones, but their nationality is difficult to determine, says Yamazaki. Although researchers would very much like to resolve both the controversy and their dispute with local residents before moving in, says Yamazaki, they believe that the dangerous condition of their present building requires them to move immediately.

David Swinbanks

NEWS IN BRIEF

Washington. Igor Shafarevich, a Russian mathematician who was last month asked to resign from the US National Academy of Sciences for alleged antisemitic writings and actions (see *Nature* 358, 442; 1992), has told the academy that he will not do so. But a US mathematician who led the unprecedented campaign to oust Shafarevich remains hopeful that the academy's actions will induce Shafarevich's employer, the Steklov Institute in Moscow, to improve its record of hiring and promoting Jewish researchers. In a letter dated 4 August, Shafarevich called the accusations of Frank Press, academy president, "absurd and scandalous", adding that "I do not know how you could have found positions in my work which do not exist". Although Press says that the academy cannot force Shafarevich to resign, Lawrence Shepp of AT&T Bell Laboratories predicts that the Steklov Institute may quietly decide to adopt a policy of nondiscrimination during next month's election of a new director. **J.M.**

Washington. The US National Science Foundation has awarded grants to four universities in the second round of a programme that makes states, industry and the federal government equal partners in a cooperative research centre. Choosing from 24 applications, NSF agreed last week to give the University of Massachusetts at Amherst, the University of Oklahoma, the University of Rochester, New York, and the University of Texas at San Antonio roughly \$1.2 million over the next four years to support basic research related to various industrial applications. The programme is modelled on one begun more than a decade ago in which the federal government provides seed money for a centre that eventually derives its support solely from the private sector. NSF made six grants in the first year of the programme, developed in conjunction with the National Governors Association, but has no plans as yet for a third round because of an expected level budget in 1993. **J.M.**

Research groups to leave Hunterian Institute

London. Basic research at the Royal College of Surgeons of England is coming to the end during the next several months as its research departments are either moved out or closed down. Although the college says that it will continue to fund the groups, the affected departments are trading a secure home for the unstable world of the London teaching hospitals.

The changes come as a result of an internal review of the workings of the Hunterian Institute, that part of the college responsible for research and education. The college's president, Norman Browse, declined to comment on the changes, but officials are said to feel that the college is no longer the best place to carry out basic research and that the departments would be better off closer to their clinical counterparts.

Outside observers say that space has become a problem, with laboratories being swallowed up by administrative offices. Some say that a lack of space has prevented the research groups from attaining or maintaining the critical mass they need to be really successful. What will remain after the changes, apart from administrative offices, is an expanded anatomy department devoted to applied surgical research and a greater emphasis on teaching.

The institute was formed in 1986 by the merger of the Institute of Basic Medical Science, co-administered by the college and the University of London, and the college's own research departments in anaesthetics, ophthalmology and the dental sciences. In 1990–91 the Hunterian Institute spent £2.3 million (US\$4.5 million) on research, drawn from both the college and external project grants.

Most of the changes are taking place within the departments from the old Institute of Basic Medical Science. Biophysics and pharmacology will both be moving, while a decision about pathology is awaited. The biochemistry department will be closed when its head, Neville Crawford, retires. The master of the Hunterian Institute, Sir Stanley Peart, will also

be retiring later this year, and the college has yet to decide whether to appoint a successor.

Although the Hunterian itself is only six years old, the tradition of research at the college dates from the 1950s and includes such prominent figures as Nobel laureate Sir John Vane. The Imperial Cancer Research Fund (ICRF) laboratories at Lincoln's Inn Fields began as part of the college before the need for more space forced it to move next door. Links between the two are still strong, with some sharing of library and scientific resources, and staff at the ICRF are unhappy with the move.

Meanwhile, the groups' new landlords are not overwhelmingly inviting. Medical researchers in London are awaiting next month's publication of the Tomlinson re-

port, which is expected to suggest a strategic rearrangement of the London teaching hospitals, some closures and an inevitable loss of research teams. Some claim that the uncertainty alone has damaged research work (see *Nature* 357, 617; 1992).

Although staff at the institute are unhappy about the changes, they are reluctant to speak publicly. They say that they expect the college to continue its support and that their research groups are "fit and well". However, they are more outspoken on the subject of the general malaise in medical research. "Wherever you look, all you can see is an 'if'," says Nina Wedderburn, academic dean of the institute and a member of its pathology department.

Ian Mundell

He's not heavy, he's my fossil

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

A Chinook helicopter lifts a rock and plaster bundle containing most of a stegosaurus, a manoeuvre required to extract the rare specimen from the Colorado site where it was found in June. The nearly complete stegosaurus skeleton includes only the second intact skull to have been found, according to Ken Carpenter, a Denver Museum of Natural History palaeontologist who led the dig.

Because the skeleton is so well-preserved, researchers have been able to confirm that the stegosaurus's characteristic back plates are arranged in an alternating pattern (some theories had them in matched pairs) and that the animal had even more body armour than had been previously thought, including a disk-shaped plate near its hip and a web of ossicles — small coin-sized bony plates — in its throat region. Carpenter notes that the new find's back plates are much larger than a previous specimen (presumably of the same species) found only 300 yards away in 1931, which may suggest that plate size is an indication of gender and that the specimen is a male.

Eric Small/DMNH

A national primate centre?

SIR — Your plan to keep the London Zoo going (*Nature* 358, 354; 1992) is misplaced; the British government and the zoo's potential customers have made it clear that they will not support the zoo. And, indeed, its closure in its present form would be no great tragedy. It cannot compete with wildlife parks, many within easy reach of London. It would be a tragedy if the Zoological Society became the steward of a disused wreck.

On the other hand, the premises at the zoo could be put to good use by turning them into a national primate centre. At present, research on primates is carried out in several different laboratories across the United Kingdom. The research is expensive and, because of the way these laboratories are dispersed, largely inefficient. At least in the centres I have visited, the conditions under which primates are kept are not ideal, and the costs of important new initiatives in primate research, such as neuro-imaging, are prohibitive.

All that would change if such research were concentrated in one central national facility catering for all fields requiring primate research — studies of higher brain function, of neurodegenerative disorders and of AIDS are obvious examples. For example, the ideal facility for the study of higher brain function and neurodegenerative disorders would include positron and/or single photon emission tomography and a 1.5 T magnetic resonance imager, as well as high-grade neurophysiological equipment and first-rate computing facilities. No British university is likely to be able to establish such a facility, which is why a national primate centre, at what is now the London Zoo, is a compelling need.

After all, the site already has space, some facilities and the expertise of the Zoological Society and the staff to breed primates humanely. Indeed, the arrangement might dispense with the need to import primates. Funding for the centre would be provided by the Zoological Society with appropriate contributions from research councils and perhaps the pharmaceutical industry. The centre could provide excellent training facilities and even the most pessimistic calculation would show that the cost of concentrating Britain's primate research in one centre would be substantially less than the present cost of funding primate research at individual laboratories.

There would be many other advantages. The safety and well-being of the animals and staff could be more easily secured than at present. Indeed, the scale of the operation would probably justify a resident Home Office inspector,

while the expertise of the Zoological Society and of the zoo staff could ensure that primate research carried out at the centre was absolutely necessary and justified and that conditions were as near as possible to the ideal.

A centre on this scale, run on the research hotel principle, could indeed be of immense scientific importance, bringing together as it would perhaps 200 or so scientists and support staff from different parts of Britain and elsewhere who would be working together on problems ranging across the whole field of biomedical science. And this development would at a stroke free the Zoological Society from its Victorian shackles and allow it a central role in contemporary biological and medical research.

George Fink

MRC Brain Metabolism Unit,
University Department of Pharmacology,
1 George Square,
Edinburgh EH8 9JZ, UK

Sheep count

SIR — Someone who, like Henry Gee (*Nature* 357, 639; 1992), seeks "to conserve the greatest amount of biodiversity in the smallest area" has missed the point of conservation. As Gee himself says, "convincing arguments stem from the concept of biodiversity as the sum of the interactions between species". Humankind is one of those species, part of nature, not set apart from it. Anything that diminishes nature diminishes humankind. To avoid that, we need to conserve the greatest amount of biodiversity in the largest area.

What is this life if, hooked on economic growth, there are no sheep and cows and other phenomena of nature to stand and stare at?

Ian Davidson

Holly Wood House,
Broom Way, Oatlands Park,
Weybridge, Surrey KT13 9TG, UK

AT&T research

SIR — The article on AT&T Bell Laboratories' research (*Nature* 356, 184; 1992) suggesting that we are drastically cutting back on research misrepresents reality.

In 1980, the Bell Laboratories' budget for research and development was \$1.25 billion; in 1991, it was approximately \$3.1 billion. About 10 per cent of the total budget continues to be allocated to basic research.

Christopher Anderson also misunderstands my use of the ratio 60:40, which was a rough figure for the ratio of

research in physical and materials sciences against software sciences. As vice-president for research Arno Penzias has stated, we are planning to make this ratio closer to 45:55. The ratio, of course, has nothing to do with the proportion of our efforts devoted to basic research.

Anderson also quotes me as saying that we are abandoning basic research on superconducting materials in favour of the development of superconducting wires. In reality, we have reduced or eliminated our effort on wires and are concentrating on the basic properties of the materials.

On Anderson's more general point about the future of basic research at Bell Laboratories, I have no reason to believe that the laboratories will ever pull out of basic research.

C. K. N. Patel

AT&T Bell Laboratories,
Room 1A-222, 600 Mountain Avenue,
Murray Hill, New Jersey 07974-0636, USA

Catching AIDS

SIR — Is it really 1992? The question posed by Lee Frank of the University of Miami School of Medicine (*Nature* 358, 10; 1992) is annoying at best and frightening to say the least: "I receive a relatively large number of reprints from areas of the world where AIDS is rampant. . . would like assurance. . . if one gets a paper-cut finger from the saliva-licked envelope. . . needn't worry about the possibility of viable AIDS virus passage."

I am not a scientist, researcher or doctor but I am HIV positive and see myself as a "knowledgeable reader". It has been stated many times that the HIV virus is an extremely difficult virus to "catch"; that the amounts of virus found in saliva and other bodily fluids (other than blood, semen and vaginal fluids) are very small; and that the protein envelope is extremely fragile and cannot survive outside the body for any length of time. Given the number of infected people in the United States and the world, surely there would be very few individuals left uninfected if the virus were so easily transmitted. One needs to go only as far as the nearest library, local AIDS organization or physician for information. Or what about reading *Time* magazine? I think I can assure Mr Frank that he is safe. I would like to know if he has received any reprints concerning AIDS. If so, it is time to start reading them.

William Truett

University of Southern California,
Graduate Program in Molecular Biology,
Los Angeles,
California 90089-1340, USA

Air pollutants and lung cancer

SIR — Simon P. Wolff's endorsement¹ of Paoletti *et al.*'s research programme is seconded, but the remainder of his letter may confuse your readers, either because of neglect of the importance of latency in the development of lung cancer or because of inappropriate generalization.

On the basis of what is known of the major contributions of radiation, chemical exposure and tobacco smoking, cases of lung cancer occurring today reflect the respiratory exposures of 20 to 30 years ago². This latency is commonly observed for many types of cancer. The mix of air pollutants 20–30 years ago in Britain was largely derived from household heating, industrial emissions and power generation. Urban rates were and are higher than rural ones, but in Britain, the measured levels of pollution did not correlate well with local cancer rates, although such rates were correlated well with population density. In the United States, lung cancer rates were higher in medium-sized towns than in large cities, and for recent immigrants to towns and cities than for lifetime residents. Such findings led to the general conclusion that the urban excess of lung cancer was not likely to be due to air pollution³.

In California, where I was in charge of research on the health effects of air pollution for the State Health Department, we demonstrated quite effectively that motor vehicle pollutants were a health risk, and took the lead in abating such effects⁴, but we did not find evidence for such pollutants being a cancer risk, with or without taking latency into account.

I know of no epidemiologists who lay the entire blame for lung cancer at the door of smoking, although most agree that it is the most important single cause. I feel that the evidence is somewhat weaker than does Doll, but subscribe to his current assessment of the urban excess of lung cancer: "In many towns coal smoke has now been replaced by the exhaust fumes of cars, which may be much more irritating, but are less likely to cause cancer, if we can judge by the quantity of polycyclic hydrocarbons they emit"⁵. Polish studies still show coal-based pollution associated with lung cancer⁶.

I agree with Wolff that more research is justified, but, in view of the latency problem, cross-sectional studies asking how many vehicles pass a place of residence would not be useful, because the types and magnitude of vehicular emissions must surely have changed over the decades. As for smoking, a lifetime smoking history would, it follows, be preferable to asking "How many

cigarettes do you smoke each day?".

John R. Goldsmith

*Epidemiology and Health Evaluation Unit,
Ben-Gurion University of the Negev,
Beer Sheva 84 105, PO Box 653, Israel*

1. Wolff, Simon P. *Nature* **356**, 471 (1992).
2. Whittemore, A. S. *Am. J. Epidemiol.* **106**, 418–432 (1977).
3. Goldsmith, J. R. in *Air Pollution: a Comprehensive Treatise* 3rd Edn, Vol VI, 391–463 (ed. Stern, A. C.) (Academic, New York, 1986).
4. Goldsmith, J. R. *Publ. Hlth. Rev.* **18**, 183–236 (1992).
5. Doll, R. *Int. J. Cancer* **47**, 803–810 (1991).
6. Jedrychowski, W., Becher, H., Wahrendorf, J. & Basa-Cierpielek, Z. *J. Epidemiol. Community Hlth.* **44**, 114–120 (1990).

Electoral ploy?

SIR — Alison Abbott reports (*Nature* **358**, 9; 1992) that the new Italian Minister for University and Scientific Research, Sandro Fontana, will soon be confronted with a proposal from CNR president Luigi Rossi-Bernardi to change the present rules governing the financing of research. The present system is indeed to be rejected for the very reasons indicated by Rossi-Bernardi: it does not know how to say no, thus giving small grants to most applicants but large grants to none.

Nevertheless, the article in *Nature* could leave the false impression that Rossi-Bernardi has so far somehow been prevented from adopting a new science funding strategy. In fact, he has been CNR president for the past seven years. Soon after his nomination, he identified "the key issue as that of quality: the proper allocation of resources, posts and promotions on the basis of scientific merit rather than custom or seniority" (see *Nature* **311**, 501; 1984).

The funding system could have been easily changed, if the CNR authorities wished to do so, by introducing some simple rules such as those adopted by other Italian funding agencies: site visits, an international board of reviewers and appointment of members of the scientific councils and of directors of 'subprojects' of national target projects among scientists with unquestioned scientific credentials, rather than on the basis of political or academic partnership (as often happened). Rossi-Bernardi could have sought qualified international advice for the creation and funding of CNR institutes and could in particular have allowed mobility of researchers to achieve better work. In fact, there was no compelling reason for not doing so, as many, if not all, of these goals could have been achieved without asking permission of anybody. Furthermore, there is no doubt that Rossi-Bernardi would have enjoyed the full support of the previous

minister of University and Scientific Research, Antonio Ruberti. His recent declaration, at the end of his present appointment as president of the CNR, sounds more like a request for reappointment than a statement of intent to fulfil these goals.

Paolo Amati

*Universita di Roma La Sapienza,
Dipartimento di Biopatologia Umana,
V. le Regina Elena 324,
00161 Roma, Italy*

Indian bureaucracy

SIR — In "Tough times ahead" (*Nature* **356**, 97; 1992), K. S. Jayaraman refers to the Indian government's plan to close the National Buildings Organisation (NBO), a leading research and development organization engaged since 1954 in the promotion of low-cost housing technology. It has the status of a government department under the Ministry of Urban Development, and serves as the UN Regional Centre for Housing for the ESCAP (Economic and Social Commission for Asia and the Pacific) region. The plan to close this important organization is also a saga of humiliation and subjugation of Indian scientists and engineers by the all-powerful Indian Administrative Service (IAS).

The innovative cost-saving construction techniques and building materials propagated by NBO and its 15 Regional Housing Development Centres have found large-scale application in housing projects undertaken by various construction departments and agencies and have been widely praised. The July/August 1991 issue of *Building Research and Information* says they are "worthy of replication in Asia, Africa and Latin America, regions with similar problems of poverty, large population and acute housing shortage".

Despite its record over the past 38 years, NBO has recently become a victim of neglect and apathy on the part of the IAS. It has been without a director since the previous director retired four years ago. For several years, vacant posts have not been filled. The acting director of NBO for the past two years is a senior IAS officer, whose role seems to be to abolish the NBO. Ironically, the government is at the same time promoting the interests of the new Building Materials and Technology Promotion Council (BMTPC), set up in July 1990 under the aegis of the Ministry of Urban Development, and whose budget has been enormously increased, in spite of the government's curbs on expenditure. Why not disband that instead?

A. K. Lal

*National Buildings Organisation,
Nirman Bhavan, New Delhi 110011, India*

Energy levels by path integration

After a delay of more than 40 years, it seems that a gloss on Feynman's way of doing quantum mechanics will help in calculating atomic ground states.

WHEN the late Mark Kacs joined the Rockefeller Institute (now University) in the early 1960s, not all his new colleagues were as delighted as they should have been. Rumours (never confirmed) that Kacs's salary was roughly what that of a Harvard professor would be in 1992 led some to grumble that the institution was being run as if it were a baseball team. The more durable alarm was that the arrival of a mathematician of such stature made it seem possible that the resources of the university would be spread more widely than had previously been the case. (George Uhlenbeck, a physicist, was already there.) But Kacs quickly charmed his critics by his wit and versatility. His most endearing trait was his interest in other people's problems (not a common property of mathematicians).

One seed of that charitable frame of mind has now borne fruit, surprisingly in a new way of calculating the energy of the ground-states of atoms with more than one electron. The tale is that Kacs, moved by claims that Feynman's account of quantum mechanics was not mathematically rigorous, set out to rescue that appealing way of looking at things from the opprobrium it seemed destined to attract. If Kacs had published his solution more visibly than in the *Journal of Research of the National Bureau of Standards*, it would no doubt have had a more immediate impact.

The point at issue lies at the heart of Feynman's way of doing quantum mechanics, and of the quantum electrodynamics that followed from it. Start with Young's slit experiment in quantum mechanics: a sheet of impenetrable material with two parallel slits, a source of electrons on one side and a detector on the other, yielding an array of interference fringes just as in the original optical experiment. The interference fringes are quite distinct from the spot or the line-like pattern that would be produced if one or other slit is blanked off. Bohr's explanation of this is that, with two slits, electrons do not go through either one slit or the other, but through both. In other words, the two paths through the two slits contribute to the interference pattern at the detector according to the phase they carry when they reach there.

Feynman's generalization (forgetting about slits) of this argument to the motion of any quantum system, say a free electron moving in an electric potential, was to posit that all possible parts contribute to the evolution of the system. Instead of simply adding together two numbers (the contributions to the end result of the two slits), it was

necessary to add together the outcomes of the infinite number of all possible paths, implying not summation but integration of some kind.

The underlying difficulty with Feynman's approach was that some of the possible paths from, say, *A* to *B* would inevitably be very strange. Along some paths, for example, a particle would change direction irregularly and frequently, and these paths would have to be given as much attention as more well ordered ways of making the transition. The trouble is that there must be even more irregular than well ordered paths — a multiply infinite rather than simply infinite number. How, in these circumstances, could people hope to rely on the results of some integration of the effects of all the separate paths? Kacs's resolution was to adapt to Feynman's problem a definition of path integration originally due to Wiener. The pure mathematicians were content and went back to what they had previously been doing.

How does this recondite argument lead to new solutions to the problem of the energy levels of atoms with several electrons? Kacs's argument seems to have what is called heuristic value as well as that of answering the purists. A. Korzeniowski, J. L. Fry, D. E. Orr and N. G. Fazleev from the University of Texas at Arlington have also seen it as a way of carrying through a computational scheme using a CM-2 model of the parallel computer called the Connection Machine (*Phys. Rev. Lett.* **69**, 893; 10 August 1992).

The difficulties of calculating the properties of many electron atoms have been known since the early days of quantum mechanics. At the outset, the solution of the hydrogen atom was a triumph, but then helium proved intractable. Those who would tackle such problems by solving the Schrödinger equation have a formidable task. Because the electrons repel each other, the simple wave functions describing the states of a hydrogen atom are no longer appropriate. Ideally, the states of the electrons in, say, helium could be represented as an infinite sum of all possible states of hydrogen-like atoms (with the nuclear charge doubled), and the repulsion between the electrons counted as a perturbation. The arithmetic, not to mention the algebra, is formidable.

But there is a further complication. The only allowable states of the two electrons of a helium atom are those in which the wave-function is antisymmetric with respect to the interchange of the two electrons, in accordance with Pauli's exclusion principle. The result is that people are required to calculate

with wave-functions consisting of determinants with as many rows and columns as there are electrons and whose elements are single-electron hydrogen-like wave-functions. It is not surprising that the past half-century has seen the emergence of simpler ways.

The connection between Kacs' argument and parallel computers is necessarily more recent. Its essence is Kacs's proof that the lowest energy of a quantum system is linked with the expression for the position of a particle undergoing a random walk, and entails the integration of that expression along the length of some paths accessible to the system (which makes the energy of the ground-state not a function of, but a functional of, the expression.) What can be more appropriate than a parallel computer for calculating quantities like this? Dealing with random walks by computer is simply a matter of simulating them time and time again.

Indeed, a further refinement is possible. While the handling of the Slater determinants whose elements are one-electron wave-functions leads to endless algebraic complications in the computation of the lowest energy state, following Kacs's method it is possible to exclude the states of unwanted symmetry — those which are not antisymmetric on the exchange of any pair of electrons. It is simply necessary to specify that a random walk reaching such a state is promptly terminated and excluded from the calculation.

That sounds a little inelegant as does the technology. The parallel computer embodies some thousands of microprocessors, small groups of which are assigned to the parallel integration of the mathematical expressions along particular feasible paths, themselves random walks. But the outcome is impressive. The calculated ground-state energies of atoms from hydrogen to boron agree within the statistical errors of the computation with the measured values, which are themselves accurate to a fraction of a per cent.

The advantage of this computation method is not merely that it seems to work well but that it has the virtue of not requiring in advance guesses at the kinds of wave-functions that would make the most efficient starting points for computation. And although it may be argued that a strictly computational technique for the properties of atoms robs the process of clarity, that is how these calculations finish up in any case. Kacs would have been delighted that his work, more than 40 years after the event, promises to be so useful.

John Maddox

A welcome animal model

Francis S. Collins and James M. Wilson

THE creation of mouse models for human disease by homologous recombination in embryonic stem cells has emerged as a major tool for studying molecular pathogenesis, but unfortunately these mouse mutants do not always resemble their human counterparts as closely as one would like. Two recent examples are the *hprt*⁻ mouse^{1,2}, which lacks the enzyme hypoxanthine phosphoribosyl transferase and was constructed as a model for Lesch-Nyhan syndrome, but which has no detectable phenotype, and the p53 mutant mouse³, whose phenotype is much less dramatic than would be expected from the homozygous obliteration of this tumour-suppressor gene. It was therefore with considerable anxiety and anticipation that cystic fibrosis researchers awaited the construction of a mouse model of this common and potentially lethal autosomal recessive disease. The results, reported in a pair of papers in the latest issue of *Science*^{4,5}, were well worth the wait.

Since its original clinical characterization, the unifying pathological feature of cystic fibrosis has been the obstruction of the respiratory, gastrointestinal and reproductive tracts, particularly their serous and mucous glands, by thick sticky secretions⁶. The identification of the cystic fibrosis transmembrane conductance regulator (CFTR) gene by a positional cloning strategy⁷⁻⁹ has done much to clarify the basic cellular defect, which involves a loss of the normal cyclic AMP-mediated apical chloride conductance of a wide variety of epithelial cells (for a review, see ref. 10). Presumably the resulting imbalance of chloride and water flux results in the dehydration of secretions characteristic of cystic fibrosis (CF), and is subsequently worsened by obstructive damage to involved organs, and by bacterial superinfection in the respiratory tree. But the absence of an animal model has prevented a better understanding of the pathophysiology of CF at the organ and whole animal level, and created significant obstacles to developing and testing new therapies.

Successful targeting of the murine CFTR locus turned out to be a formidable task, and has occupied the intense attention of several experienced groups over the past three years. On the way to achieving this goal, the University of North Carolina group, led by Bev Koller and Oliver Smithies⁴, had to develop a series of elegant and rigorous controls, and discovered that the frequency of homologous versus non-homologous recombinant events was capable of tenfold variations from day to day; indeed, the

identification of the factors responsible for this variability would be of enormous benefit to such experiments in the future. Eventually two cell lines harbouring an interruption of exon 10 of the CFTR gene and capable of germ-line incorporation were constructed, heterozygotes were generated, and these were bred to produce homozygous (-/-) animals.

In the light of the physiological differences between humans and mice, and in the face of previous failure to identify any spontaneous mouse model of CF, there was concern that the (-/-) animals might have an embryonic lethal phenotype. Gratifyingly, the (-/-) animals appeared in the expected mendelian ratio at birth, indicating no apparent intrauterine loss. Shortly after birth, however, these animals began dying, and few survived beyond 40 days. The cause of death appeared in all cases to be intestinal obstruction, perforation and fatal peritonitis. This part of the mouse CF phenotype bears striking similarity to the human syndrome of meconium ileus, which occurs in 5-10 per cent of CF newborns and can be fatal. Histological examination of the (-/-) mice revealed distention of the crypts of Lieberkuhn by dense concretions throughout the intestinal tract.

Physiological studies of the (-/-) mice⁵ also confirmed similarity to the human phenotype at the cellular level. In both intestinal and respiratory epithelia, Ric Boucher and colleagues demonstrated the absence of the normal cyclic AMP-stimulated apical chloride channel, which is the physiological hallmark of CF in man. As in man, murine respiratory epithelium contains an alternative non-CFTR chloride channel which responds to externally applied ATP or UTP, and which provides an attractive pharmacological alternative for the treatment of CF.

Similarity between the pathology of the mouse model and the human disease, however, did not extend as well to the reproductive, pancreatic, hepatobiliary and respiratory systems. In the reproductive tract there was no evidence for abnormality in the mouse, and one long-surviving (-/-) male successfully reproduced. Similarly, there was no conclusive evidence for pancreatic or hepatobiliary disease except for inflammation of the gall bladder. In the lungs, which are responsible for 95 per cent of the morbidity and mortality of CF in man, no evidence for obstructive disease was found in the (-/-) mice, although there was pathology in the

nasal sinuses and an increased frequency of patches of goblet cells in the proximal airway. This general lack of a pulmonary phenotype may reflect anatomical differences: rodents essentially lack the components of the human airway primarily responsible for mucous production, including submucosal glands, which are the predominant site of CFTR expression in human bronchi (J. Englehardt and J. M. W., unpublished results). The absence of clear pathological changes in the lungs, liver and pancreas of the (-/-) mice might be viewed as a disappointment. Given that such changes are quite limited in CF newborns, however, the possibility remains that these murine manifestations might have appeared had the animals survived longer.

The consequences of the availability of such a mouse model are likely to be many. The affected mice provide a golden opportunity to study the intestinal pathophysiology of CF in much greater detail than before. It will be of interest, for example, to determine whether heterozygotes are selectively resistant to intestinal exposure to cholera toxin, an advantage that might explain the high frequency of CF in the Caucasian population¹¹. To address this and other questions about CF, it may also be necessary to produce a mouse homozygous for the $\Delta F508$ mutation, which is a deletion of a phenylalanine residue at position 508 of the CFTR and is responsible for 70 per cent of CF cases in northern Europeans. This mouse might have a subtly different phenotype in both heterozygotes and homozygotes from a CFTR 'knockout' mouse. The recent development of 'hit-and-run' targeting protocols^{12,13} should now make such a precise alteration of the murine CFTR gene possible.

The CF mouse will immediately find a role in the testing of new therapies. Although the absence of pulmonary manifestations and the (potentially related) discovery of a murine airway amiloride-sensitive sodium transport system with no human homologue⁵ will

- Hooper, M., Hardy, K., Handyside, A., Hunter, S. & Monk, M. *Nature* **326**, 292-294 (1987).
- Kuehn, M. R., Bradley, A., Robertson, E. J. & Evans, M. J. *Nature* **326**, 295-298 (1987).
- Donehower, L. A. et al. *Nature* **356**, 215-221 (1992).
- Snouwaert, J. N. et al. *Science* **257**, 1083-1088 (1992).
- Clarke, L. L. et al. *Science* **257**, 1125-1128 (1992).
- Boat, T. F., Welsh, M. J. & Beaudet, A. L. in *The Metabolic Basis of Inherited Disease* (eds Scriver, C., Beaudet, A., Sly, W. & Valle, D.) 2649-2680 (McGraw-Hill, New York, 1989).
- Rommens, J. M. et al. *Science* **245**, 1059-1065 (1989).
- Riordan, J. R. et al. *Science* **245**, 1066-1073 (1989).
- Kerem, B. et al. *Science* **245**, 1073-1080 (1989).
- Collins, F. S. *Science* **256**, 774-779 (1992).
- Quinton, P. M. *Pediatr. Res.* **16**, 533-537 (1982).
- Hasty, P., Romerez-Solis, R., Krumlauf, R. & Bradley, A. *Nature* **350**, 243-246 (1991).
- Valancius, V. & Smithies, O. *Molec. cell. Biol.* **11**, 1402-1408 (1991).
- Gelehrter, T. D. & Collins, F. S. *Principles of Medical Genetics* 125 (Williams and Wilkins, Baltimore, 1990).

significantly complicate the use of this model for evaluation of pharmacological approaches to CF lung disease, there are likely to be correlations between approaches that ameliorate the intestinal phenotype in the CF mouse and those that would be beneficial in the lungs in humans. For gene therapy protocols, issues of efficiency and safety of CFTR gene transfer can already be addressed in non-CF animal models. But a genetically deficient animal that models the clinical manifestations of CF, especially in the lungs, will be necessary to answer critical questions of therapeutic efficacy; this information is fundamental to the design of early clinical protocols. It will therefore be important to transfer the existing mutant onto a wide variety of mouse strain backgrounds, and also to breed long survivors from such crosses selectively. The identification of (—) animals with milder intestinal manifestations permitting long survival may allow emergence of a pulmonary phenotype,

especially if these animals are exposed to pathogens. Such breeding experiments would also provide the opportunity to map modifier genes that determine the severity of the disease, which may well have human homologues.

Geneticists are fond of quoting the old (and perhaps somewhat self-serving) maxim: "whatever it is you're studying, you're better off if you have a mutant"¹⁴. The successful development of a mouse model for CF will undoubtedly have significant repercussions in the rigorous search for a better understanding and improved treatments for this debilitating disease, and stands as a significant milestone in the midst of a flurry of recent stunning advances in CF research. □

Francis S. Collins and James M. Wilson are in the Howard Hughes Medical Institute, the Division of Molecular Medicine, and the Cystic Fibrosis RDP Center at the University of Michigan, 4570 MSRB II, Box 0650, Ann Arbor, Michigan 48109, USA.

CYSTIC FIBROSIS

Another protein out in the cold

John Armstrong

A SINGLE amino-acid deletion in the cystic fibrosis transmembrane conductance regulator (CFTR) is the mutation most frequently responsible for cystic fibrosis. How does this change affect the normal function of the protein? The mouse lacking functional CFTR which is described above by Collins and Wilson is already on track to provide some of the answers. And a report by Denning *et al.* on page 761 of this issue¹ rounds off a series of studies indicating that the problem lies not so much with what the mutant protein cannot do as with where it ends up in the cell. Like many of the altered membrane proteins constructed by cell biologists, the mutant CFTR is unable to leave the endoplasmic reticulum, but in the best traditions of classical genetics it is temperature-sensitive: at lower temperatures the protein travels onwards to the cell surface where it can then function, more or less.

Cystic fibrosis is one of the most common human genetic diseases. The story of the identification of the gene responsible has been widely reported as a triumph of the new genetics. Just as impressive has been the progress made in understanding what this piece of DNA encodes: CFTR appears to function as a regulated channel for chloride ions². The most frequently occurring mutation, $\Delta F508$, which causes disease in homozygotes, lacks a phenylalanine residue in the first nucleotide-binding domain of the protein. The simplest result of

changing the structure of a protein is to damage its biochemical activity, but for a membrane protein there are other possibilities. CFTR must reach the plasma membrane to function. It does this in the usual way, by first assembling in the endoplasmic reticulum, then travelling via a series of vesicles through the Golgi complex to the cell surface. Failure to accomplish any of these steps will have consequences just as drastic as any loss of biochemical function.

When the faulty CFTR gene is transfected into cultured cells, much of the $\Delta F508$ protein is trapped in the endoplasmic reticulum³. In contrast, when the defective protein is expressed in an alternative system, the *Xenopus* oocyte, it appears to travel normally to the cell surface, but is not correctly regulated by cyclic AMP⁴. This raised the possibility that the apparent accumulation of mutant CFTR in cultured cells is an artefact of the expression system. Two complementary papers that argue against this explanation both rely on reagents sufficiently sensitive to detect the protein at its normal expression level. In primary cultures of normal airway epithelial cells, CFTR was found largely at the apical surface, whereas in cells derived from $\Delta F508$ patients it was inside the cell⁵. Likewise, direct antibody labelling of tissue samples from patients showed a lack of apical membrane protein in some tissues⁶.

The latest work¹ neatly resolves this

apparent discrepancy. It transpires that mutant CFTR is not absolutely blocked in the endoplasmic reticulum: when transfected cells are incubated at temperatures lower than 37 °C, the protein starts to appear at the surface. At the temperature used to incubate *Xenopus* oocytes, a significant proportion of the protein is transported correctly. Once at the surface, the behaviour of the mutant protein as a chloride channel is not completely normal, but nevertheless it retains enough of its function to imply that the main problem at body temperature is a failure of delivery to the plasma membrane.

Such behaviour is by no means unprecedented among membrane proteins. A temperature-sensitive mutant of vesicular stomatitis virus, tsO45, fails to bud at high temperature because its membrane protein, the G protein, is trapped in the endoplasmic reticulum⁷. The reversibility of this trapping has made tsO45 a favourite work-horse in studies of membrane traffic, because G protein can be accumulated in large amounts in the endoplasmic reticulum, then, by lowering the temperature, released as a 'pulse' through the secretory pathway to the surface. Some artificially mutated membrane proteins have been shown to behave similarly⁸; many more are trapped in the endoplasmic reticulum and never leave. Perhaps most remarkably, HLA class I proteins normally require binding of a peptide to leave the endoplasmic reticulum, but at lower temperatures they can escape and arrive empty-handed at the surface⁹.

The mechanism responsible for retaining such proteins in the endoplasmic reticulum is less than clear. Some but not all of them are associated with BiP, a 'chaperone' protein whose normal function is probably to regulate folding and assembly¹⁰. Simple aggregation in the membrane could be enough to prevent these molecules from entering the regions of the membrane where transport vesicles bud, a 'quality control' process that may prevent misfolded proteins from reaching the cell surface¹¹. Thermal instability, leading to unfolding and precipitation, is a frequent characteristic of mutant proteins. However, mutant CFTR can function as a channel at 35 °C, arguing that it cannot be grossly denatured¹. In addition, it is not obvious that denaturation of a membrane protein, which can move only within the plane of the lipid bilayer, would cause it to aggregate. A satisfactory description of this two-dimensional process may not come readily from biochemists trained exclusively in three dimensions.

Thus cystic fibrosis appears to have joined the list of inherited human 'protein targeting' diseases, which between them illustrate several basic concepts. A

celebrated example is the familial hypercholesterolaemia caused by mutations in the low-density lipoprotein receptor: one particular mutant protein binds its ligand at the cell surface but fails to enter clathrin-coated pits, and hence is not taken into the cell¹². I-cell disease is an inborn error of metabolism resulting in failure to attach mannose-6-phosphate to glycoproteins: this is the signal responsible for diverting lysosomal enzymes out of the secretory pathway in the Golgi complex and sending them to lysosomes¹³. Perhaps the oddest example comes from a type of hyperoxaluria, in which the enzyme alanine-glyoxalate aminotransferase has a mutation in its targeting signal for peroxisomes. Instead of leaving the protein adrift in the cytosol, however, the mutation uncovers a 'cryptic' targeting signal for mitochondria, a bizarre case of intracellular delusion¹⁴.

This explanation for the behaviour of mutant CFTR will be very satisfying to cell biologists. How might it serve the more pressing needs of the patient? Denning *et al.* point out that the obvious treatment, cooling the patient to the point where traffic of the protein to the surface is restored, could be too drastic.

A second proposal, to interfere with the function of the endoplasmic reticulum so that the arrested protein can escape to the cell surface^{1,3}, would be an achievement to match anything so far accomplished in the field of cystic fibrosis. It is something that cell biologists would also like to know how to do. □

John Armstrong is in the Imperial Cancer Research Fund, PO Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK.

- Denning, G. M., Anderson, M. P., Marshall, J., Smith, A. & Welsh, M. J. *Nature* **358**, 761–764 (1992).
- Higgins, C. *Nature* **358**, 536 (1992).
- Cheng, S. H. *et al.* *Cell* **63**, 827–834 (1990).
- Drumm, M. L. *et al.* *Science* **254**, 1797–1799 (1991).
- Denning, G. M., Ostedgaard, L. S. & Welsh, M. J. *J. Cell Biol.* **118**, 551–559 (1992).
- Kartner, N., Augustinas, O., Jensen, T. J., Naismith, A. L. & Riordan, J. R. *Nature Genet.* **1**, 321–327 (1992).
- Knipe, D. M., Baltimore, D. & Lodish, H. F. *J. Cell Biol.* **21**, 1149–1158 (1977).
- Machamer, C. E. & Rose, J. K. *J. Biol. Chem.* **263**, 5955–5960 (1988).
- Ljunggren, H. G. *et al.* *Nature* **346**, 476–480 (1990).
- Gething, M. J. & Sambrook, J. *Nature* **355**, 33–45 (1992).
- Hurtley, S. & Helenius, A. *Rev. Cell Biol.* **5**, 277–307 (1989).
- Davis, C. G. *et al.* *Cell* **45**, 15–24 (1986).
- Kornfeld, S. & Mellman, I. *Rev. Cell Biol.* **5**, 483–525 (1989).
- Purdue, P. E., Allsop, J., Isaya, G., Rosenberg, L. E. & Danpure, C. J. *Proc. natn. Acad. Sci. U.S.A.* **88**, 10900–10904 (1991).

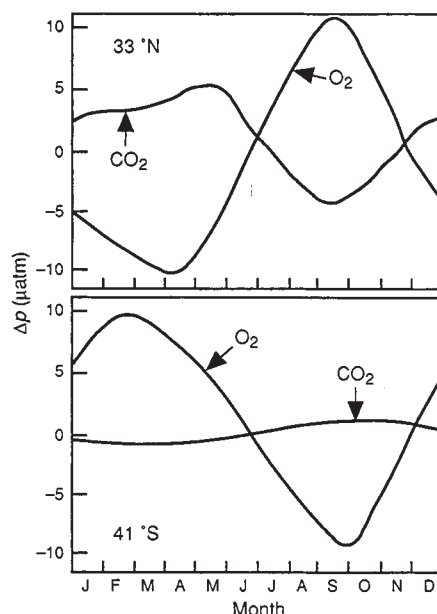


FIG. 2 Seasonal cycles in atmospheric O_2 and CO_2 for stations in the Northern and Southern Hemispheres. The magnitude of the O_2 cycles reflect the seasonality in the rate of photosynthesis both on land and at sea (high O_2 production in summer and low in winter). The magnitude of the CO_2 cycle reflects only terrestrial photosynthesis because of the sea's ability to buffer CO_2 changes.

GLOBAL CHANGE

Diminishing oxygen

Wallace S. Broecker and Jeffrey P. Severinghaus

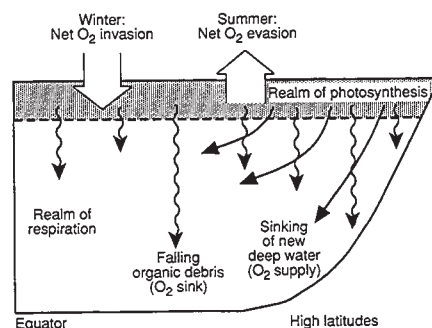
A METHOD for keeping track of changes in the atmosphere's oxygen content, described by Keeling and Shertz on page 723 of this issue¹, could make it possible to unravel the complexities of the carbon cycle, which is essential to any understanding of CO_2 variations and greenhouse warming.

The authors use an interferometric technique based on the refractive index of dry air to measure the ratio of O_2 to N_2 to within 5 parts per million. Assuming that the atmosphere's N_2 inventory

remains constant, this translates into the ability to detect single-microatmosphere changes in the partial pressure of O_2 . The authors explore two applications for this new technique. First they use it to assess the rate of decline of the atmosphere's O_2 content over the past 3 years resulting from fossil-fuel burning and deforestation. Second they use the amplitude of seasonal changes in this ratio to estimate the global production rate of plant matter in the sea.

The first of these applications is rel-

FIG. 1 The O_2 cycle for a single quadrant of the ocean. Oxygen is produced in the sunlit upper ocean by photosynthesis. The O_2 matching that portion of the photosynthetic product which sinks into the ocean's interior evades to the overlying atmosphere. However, as 99.9 per cent of the organic matter is ultimately oxidized, this O_2 is reabsorbed by the sea in areas like the equator where O_2 -depleted deep water returns to the sea surface. When averaged over an entire year, O_2 production by marine plants is balanced by O_2 consumption by marine animals and bacteria. However this balance is not achieved in a single season. During the spring and summer months, when oceanic plant productivity is at its peak, the ocean exports O_2 to the atmosphere. During the winter when the demand by upwelling deep water exceeds production by plants O_2 is imported by the sea. This creates the seasonal cycle in O_2/N_2 ratios observed by Keeling and Shertz.



evant to the long-standing problem of how much anthropogenic CO_2 is going into the ocean and how much is being reabsorbed by photosynthesis. Answering this riddle depends on knowing the rate at which the standing stock of carbon stored in terrestrial vegetation and soil humus is changing. As deforestation decreases the former and agriculture the latter, many have concluded that this combined carbon reservoir is shrinking and thereby is contributing to the ongoing rise in atmospheric CO_2 content². Yet attempts by geochemists to balance the Earth's carbon budget all conclude that the combined increase in atmospheric CO_2 content (measured) and in oceanic carbon content (calculated) fall short of accounting for the fate of the CO_2 released by the burning of coal, oil and natural gas. This leads them to the conclusion that carbon storage in the terrestrial reservoirs has increased despite the toll taken by foresters and farmers³. They point to a global greening of the Earth's natural vegetation (and soils) driven by increased atmospheric CO_2 content and by the fallout of anthropogenic NO_3 and NH_4 nutrients. But because the amount of CO_2 taken up by the ocean is calculated using imperfect models, a sizeable error plagues this budget approach.

Keeling and Shertz provide a means to circumvent this problem. The ocean contains 60 times more carbon than the atmosphere, but it contains only a few

per cent of the Earth's O_2 . Hence, although the oceanic reservoir strongly buffers atmospheric CO_2 changes, it has little impact on the atmosphere's O_2 content. Because of this the authors can directly compare the magnitude of the atmospheric O_2 decrease with that required for the combustion of coal, oil and natural gas during the 3-year period spanned by their measurements. Their conclusion is that the O_2 decrease is consistent with the fossil fuel demand: there is no need to call upon either a net production or net consumption of O_2 by the biosphere.

In other words, global greening appears to be compensating for losses by deforestation and agriculture. This implies that the ocean must be taking up all the CO_2 derived from fossil fuels that does not stay in the atmosphere, about 3 gigatonnes of carbon each year. Keeling and Shertz stress, however, that the uncertainty in their estimate of the rate of change in the amount of carbon stored in the terrestrial biosphere is as yet too large to permit firm conclusions to be drawn as to whether this reservoir is a net source or sink for CO_2 . Only when the record covers a decade will the accuracy of the trend be sufficient to make this distinction.

While waiting for their record to grow to the desired length, Keeling and Shertz have taken on a second application of their technique. It struck them that the magnitude of the seasonal cycle in the O_2/N_2 revealed by their measurements could be harnessed to place limits on the amount of plant matter which falls from the surface ocean into the dark interior (Fig. 1). Biological oceanographers call this quantity new production in order to distinguish it from the total amount of organic matter formed by photosynthesis. New production is considerably less than total production because much of the photosynthetic product is recycled by animals and bacteria living within the sunlit upper ocean. Only that portion of the organic matter which escapes the mixed layer by falling (or swimming) into the sea's interior contributes to the cycles in the sea's nutrient constituents and to the atmosphere's O_2 content.

The magnitude of new production in the ocean has in recent years become a subject of considerable debate. Geochemists using isotopic clocks⁴ estimate higher values than the traditional ones obtained by biologists using the results of ^{14}C - and ^{15}N -tagged incubations (refs 5, 6, respectively).

Keeling and Shertz again exploit the fact that the ocean's carbon system strongly buffers changes in CO_2 . The seasonal extraction of CO_2 from surface waters by marine photosynthesis is so buffered that it makes no discernible imprint on the atmosphere's CO_2 con-

Grave guardians with a difference

EVALUATING the condition of a collection of wooden figurines from the Pacific Northwest of America, Robert A. Blanchette and colleagues found that things were not quite what they seemed (*Mycologia* 84, 119–124; 1992) — many of the objects were in fact carved from the fruiting body of the wood-destroying fungus *Fomitopsis officinalis*. With hindsight, the finding might have been expected. The fungus, known to some of the native American tribes as the bread of ghosts, featured large in their culture and was attributed with supernatural powers. Indeed, the collection of the Field Museum of Natural History, Chicago, includes an uncarved fruiting body (sporophore) which had been used as a poultice to relieve swellings and inflammations. Found in coniferous forests, *F. officinalis* may have been collected by shamans ('medicine men') during retreats which they took "to encounter the spirits". The carved spirit figures now examined by Blanchette and colleagues had been collected from the late 1800s from shaman graves, where they had presumably been placed as guardians. Horizontal striations across the figures, due to the annual accumulations to the perennial sporophores, must have been mistaken for tree rings by collectors, reinforcing the incorrect conviction that the objects were fashioned from wood. Blanchette and colleagues have already identified 13 fungal carvings from five museums.

R.P.

Smithsonian Natural History Museum, Washington, DC.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

tent. They demonstrate this through a comparison of the atmospheric CO_2 and O_2 cycles in the Northern and Southern Hemispheres (Fig. 2). Because of the seasonality in terrestrial photosynthesis, the land-rich Northern Hemisphere's atmosphere shows distinct seasonal cycles for both O_2 and CO_2 . Because about half the Northern Hemisphere's photosynthesis is on land and about half in the sea the amplitude of the CO_2 cycle is about half that for the O_2 cycle. By contrast in the land-poor Southern Hemisphere, there is only a barely discernible CO_2 cycle but a large O_2 cycle. While the magnitude of the O_2 cycle has the potential to constrain estimates of new production in the ocean, Keeling and Shertz point out that optimal use of this information will require corrections for dilution through atmospheric mixing (both vertical and meridional), for the seasonal overlap between the upwelling of O_2 -deficient deep water and the O_2 supply by new production and for the contribution of terrestrial photosynthesis especially in the Northern Hemisphere.

Humanity owes much to the Keeling family. Throughout his scientific career, father Charles David has made ultra-precise measurements of CO_2 in the atmosphere and ocean which provide the base upon which our ideas regarding the fate of fossil fuel CO_2 are constructed. Now son Ralph has launched a complementary O_2 measurement programme that will probe a serious unanswered question concerning the Earth's current carbon budget. □

Wallace S. Broecker and Jeffrey P. Severinghaus are at the Lamont-Doherty Geological Observatory, Palisades, New York 10964, USA.

1. Keeling, R. F. & Shertz, S. R. *Nature* **358**, 723–727 (1992).
2. Houghton, J. T., Jenkins, G. J. & Ephraums, J. J. (eds) *Climate Change: The IPCC Scientific Assessment* (Cambridge University Press, 1990).
3. Tans, P. P., Fung, I. Y. & Takahashi, T. *Science* **247**, 1431–1438 (1990).
4. Jenkins, W. J. & Goldman, J. C. *J. mar. Res.* **43**, 465–491 (1985).
5. Eppley, R. W. & Peterson, B. J. *Nature* **282**, 677–680 (1979).
6. Dugdale, R. C. & Goering, J. J. *Limnol. Oceanogr.* **12**, 196–206 (1967).

Arithmetic in the cradle

P. E. Bryant

BABIES can apparently make simple arithmetical calculations, according to a series of experiments described by Karen Wynn on page 749 of this issue¹. The appearance of this paper is a notable event in the history of developmental psychology — Wynn claims that babies as young as five months in age are able to add and subtract, and she concludes that the basis of arithmetical understanding may be innate. Her results and her conclusion are surprising, even in an area of research which has already produced many surprises.

Over the past 20 years we have learned a great deal about the capacities of very young babies. Well within their first six months, we now know, babies can tell objects apart by their shape², size³ and colour⁴: they know that these objects go on existing when they are hidden⁵ and they also can take in whether they are solid or not⁶. Babies can even tell, when they listen to someone talking, whether that person's lip movements are the appropriate ones for the speech that they are hearing^{7,8}. These are striking skills to find in a creature which used to be thought of as utterly incompetent and ineffective, but they could all in one way or another be described — even dismissed — as perceptual. One could argue in the face of this evidence that the perceptual system of very young babies is intact, so that they can take in information about their physical surrounds, but that they may not be able to think or reason in any way.

This distinction has begun to look implausible, however, for there is now considerable evidence that children can recognize the number of a set of objects and can tell when there has been a change in that number — at any rate when small numbers are involved — and that they can do so irrespective of changes in the spatial arrangement of the set. They can even tell whether the number of drumbeats that is played to them coincides with the number of objects in a visual display that is shown to them at the same time⁹⁻¹¹. To recognize that two sets have the same number despite radical differences in their perceptual appearance is to go well beyond simple perception.

The experiments described in this issue provide remarkable evidence that young babies' intellectual skills may go a good deal further even than this. Not only do they take in information about the number of objects before them: they may also be able to work out the results

of an addition or a subtraction.

Wynn used a technique which has proved extremely useful in the past. Babies tend to look for a relatively long time at displays that are new or unexpected to them. In her first two experiments the babies saw a hand putting one or two objects on a table and then a screen being raised in front of the object(s). Then they saw the same hand either adding an object to the one behind the screen (1+1) or taking one

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Counting games: do they add up?

away (2-1) from behind it. Finally the screen was lowered, and the baby could see how many objects there were now on the table. Half the time that number of objects was as it should have been, but the other half it was not. In the latter case, the babies saw two objects when there should have been one, or one object when there should have been two.

Wynn reports that the babies looked significantly longer at these incorrect displays than at the correct ones and argues from this that they had anticipated the correct number of objects and were surprised when they saw a different number. To do so, she claims, they must have carried out some form of mental arithmetic.

Her final experiment tackles a possible objection to this conclusion. In the first two experiments the number of objects in the incorrect trials was the same as their number before the addition or subtraction: so the babies might simply have realized that the final number of objects would have to be different from the original number, and therefore were surprised after the screen was lifted to see no change. In the third experiment Wynn took care of this objection by giving them an addition (1+1) in the same way as before, but arranging things so that after the screen was lifted they either saw two objects there (the correct display) or three. Thus the number of

objects in the incorrect display was not the same as the number before the addition. Yet the babies still looked at the incorrect displays for a longer time than they looked at the correct ones.

The control is a convincing one, though it should be added that Wynn's argument would have been stronger if she had put in the same control for subtraction as well. As it is, she has made a convincing case only for babies of five months being able to perform addition. It is quite easy to think of an experiment to establish as convincing a case about subtraction.

The apparently cast-iron evidence that Wynn offers us to show that very young babies can carry out additions in their head establishes that they are capable of arithmetical reasoning very early on in life. But this prompts some further questions. One is whether young babies have a full understanding of addition and subtraction. In my view Jean Piaget was correct when he argued that someone who can add or subtract correctly does not necessarily understand addition and subtraction¹². One has to grasp the relationship between these two mathematical operations in order to understand either of them properly. Piaget claimed that children do not realize the inverse relationship between addition and subtraction until they are roughly eight years old, and only then as a result are able to learn about the additive composition of numbers (for example, that if 5 and 3 add up to 8, then 8-5 must equal 3). This is an idea that has never been challenged empirically, but Wynn's work suggests that it would be quite easy, and well worth the trouble, to devise an experiment to see whether babies are able to understand the inverse relationship between addition and subtraction.

Another very different question concerns the relationship between the early skills that Wynn has discovered in babies and the progress that they make in

1. Wynn, K. *Nature* **358**, 749-750 (1992).
2. Slater, A., Morison, V. & Rose, D. *Br. J. dev. Psychol.* **1**, 135-142 (1983).
3. Slater, A. & Morison, V. *Perception* **14**, 337-344 (1985).
4. Adams, R. J., Maurer, D. & Davis, M. J. *exp. Psychol.* **41**, 267-281 (1986).
5. Baillargeon, R., Spelke, S. & Wasserman, S. *Cognition* **20**, 191-208 (1985).
6. Baillargeon, R. *Cog. Dev.* **2**, 179-200 (1987).
7. Dodd, B. *Cog. Psychol.* **11**, 478-584 (1979).
8. Kuhl, P. K. & Meltzoff, A. N. *Science* **218**, 1138-1141 (1982).
9. Antell, S. E. & Keating, D. P. *Child Dev.* **54**, 695-701 (1983).
10. Starkey, P. & Cooper, R. *Science* **210**, 1033-1034 (1980).
11. Starkey, P., Spelke, E. S. & Gelman, R. *Cognition* **36**, 97-128 (1990).
12. Piaget, J. *The Child's Conception of Number* (Routledge & Kegan Paul, London, 1952).
13. Bornstein, M. H. & Sigman, M. D. *Child Dev.* **57**, 251-274 (1986).
14. Rose, D. & Slater, A. *Intelligence* **10**, 251-263 (1986).

mathematics later on. Wynn speculates that babies' evident arithmetical capacity "may provide the foundations for the development of future arithmetical knowledge". The suggestion may seem extravagant, but it is really quite a plausible one, given the results of some longitudinal work which shows that individual babies' performance in other experiments involving measures of looking predicts their abilities later on reasonably well. Several studies^{13,14} have established that looking time in experiments on babies' perceptual skills is strongly related to their later IQ. It is a curious relationship, given the simplicity of the earlier measures and the very general nature of the IQ test.

Wynn's hypothesis suggests that one should try to find a more specific and

therefore more satisfying connection. It seems quite possible to me that her early measures of babies' addition might be strongly related to their later success in mathematics but not, say, in learning to read. Other measures, possibly to do with sound discrimination, might predict their later success in reading but not in mathematics. Wynn's interesting speculation that she may have uncovered the foundations of arithmetical knowledge should remind us that we can no longer be content with discovering surprising skills in very young babies. We must explore the significance of these skills in children's lives later on. □

P. E. Bryant is in the Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK.

NEW MATERIALS

No contractile obligations

R. Lakes

STRETCH most materials and you will expect to see their cross-section shrink. But Alderson and Evans¹ have synthesized a microporous polymer that swells as it is stretched, and Milton² has designed a laminated composite that does the same.

These advances in the newly developing science of materials with a negative Poisson's ratio (the Poisson's ratio is the ratio of lateral contraction to longitudinal extension in the direction of stretching force) promise improved control of properties and open the door to a new class of applications. Virtually all common materials become narrower in cross-section when they are stretched. The reason, in the continuum view, is that most materials resist a change in volume more than they resist a change in shape. Specifically, Poisson's ratio is $\nu = (3K - 2G)/(6K + 2G)$ with K as the bulk modulus and G as the shear modulus.

In a structural view, a material is made of atoms interacting by central forces (springs) between nearest and next nearest neighbours. Stretching of the material will stretch the diagonal springs, causing a lateral contraction.

It is possible to have materials that become fatter when stretched because the theory of isotropic elasticity³ allows Poisson's ratios in the range from -1 to $1/2$. Physically the reason is that for the material to be stable, the stiffnesses must be positive; the bulk and shear stiffnesses are interrelated by formulae which incorporate Poisson's ratio. Familiar materials have positive Poisson's ratios, typically about $1/3$ both for stiff materials such as steel and aluminium and for compliant spongy materials, and just under $1/2$ for rubber. Because negative-Poisson's-ratio materials (or 'auxetic' materials in Evans's vocabulary) easily undergo volume changes but resist shape changes, they may be viewed as the opposite of rubbery materials.

Several examples are known. Honeycombs with inverted hexagonal cells have negative Poisson's ratios in two dimensions^{4,5} and may be viewed as structures. Some highly anisotropic materials including a few natural single crystals, some synthetic off-axis composite laminates, and microporous oriented polytetrafluoroethylene⁶ have been reported to have a negative Poisson's ratio in some

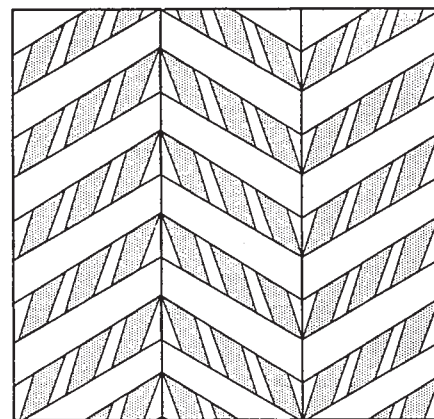


FIG. 2 Negative-Poisson's-ratio laminate of Milton. The compliant phase is shaded.

directions. Foams, based on polymers or metals, with a 're-entrant' cell structure (which bulges inward) exhibit negative Poisson's ratios in three dimensions⁷ and can be made isotropic. The foams represent the first intentional synthetic example and the first which can be made isotropic. (Further examples and the mechanisms are reviewed in ref. 8.) The honeycombs and foams need to be highly porous to achieve negative Poisson's ratios. They are therefore substantially less stiff than the solids from which they are made, and Evans *et al.* proposed that stiffer materials could be made by creating re-entrant structure on the molecular level⁹.

The new work of Alderson and Evans¹, to appear in a forthcoming issue of *Polymer*, deals with formation of microporous ultra-high molecular weight polyethylene (UHMPE) with a negative Poisson's ratio. The processing required to achieve this property consisted of compaction of powdered polymer, sintering and extrusion. The material was found following processing to contain a network of microscopic size particles connected by fibrils (Fig. 1). As a result of the extrusion process, the polyethylene was mechanically anisotropic, with Poisson's ratios as low as -1.24 for compression in the radial direction, and zero for compression in the axial direction. The degree of anisotropy was less than in the expanded polytetrafluoroethylene examined earlier⁶.

A family of laminate structures designed by Milton² allows isotropic elastic properties with Poisson's ratio approaching -1 , which is the lower limit for isotropic solids. Milton prefers to call negative-Poisson's-ratio solids 'dilatational'. The laminates have a chevron structure, with multiple length scales in two dimensions (Fig. 2); a three-dimensional version is also described. The laminates are structurally anisotropic, but can be made mechanically isotropic by proper choice of geometry and of constituent properties. ▶

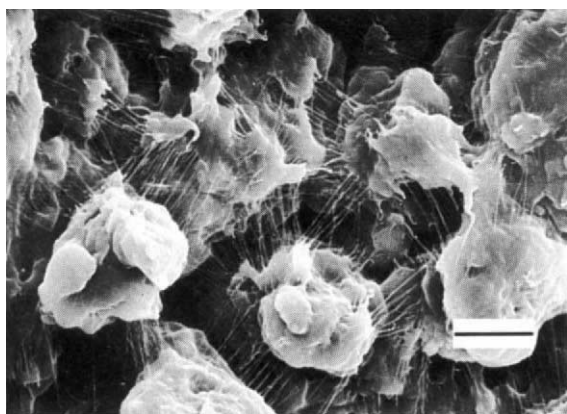


FIG. 1 Micrograph of the 'auxetic' polyethylene made by Alderson and Evans. Scale bar, 25 μm .

Surviving subduction

William M. White

The lowest Poisson's ratios of the laminates approach the rigorous lower bounds of Cherkashev and Gibiansky¹⁰. To achieve a negative Poisson's ratio in the laminate², the two constituents must differ in stiffness by more than a factor of about 25; to approach a Poisson's ratio of -1, the ratio in constituent stiffness must become large. These developments suggest that it will be possible to obtain high density and a corresponding high stiffness in materials with negative Poisson's ratios. Such materials could be of great interest in structural applications.

Work on these strange materials is not meant just for our entertainment; they may be useful for a variety of reasons. First, the Poisson's ratio itself can confer useful properties. For example, one may envisage a press-fit fastener which contracts laterally when inserted into a socket, facilitating insertion, then expands laterally to resist any attempt to remove it. In a wrestling mat, it is desirable to minimize the impact force for both small shapes (an elbow) and larger ones (a back). The lowest possible Poisson's ratio is desirable in such an application as the distribution of stress in the mat depends on Poisson's ratio. Poisson's ratio also influences the concentration of stress around holes and cracks; in some geometries a negative value could be helpful.

Second, other physical properties arise indirectly as a consequence of the unusual microstructure of negative-Poisson's-ratio materials. For example, the negative-Poisson's-ratio foams are more resilient, tougher, and more compliant than foams of conventional structure, and are effective absorbers of sound, owing to the convoluted cell ribs. In some potential applications, such as knee pads, sound-baffles, tear-resistant sponges or cores for sandwich panels, a material of low density and low stiffness is appropriate. But for structural applications a stiff material is required and until now, high stiffness has been difficult to achieve in a foam. □

R. Lakes is in the Department of Biomedical Engineering and Center for Laser Science and Engineering at the University of Iowa, Iowa City, Iowa 52242, USA.

THE continental crust has grown through geological time by the addition of material from the mantle through volcanism. But can mass be lost from the crust back to the mantle? New evidence that it can is reported on page 745 of this issue by Becker and Altherr¹. Minerals in metamorphosed limestones show textural and compositional evidence of having equilibrated at depths of 100 km or so and temperatures in excess of 1,100 °C. The rocks occur within the Variscan granulites of the Bohemian massif in the Dunkelsteiner Wald of northeast Austria and are associated in the field with peridotitic rocks of mantle origin. The limestones were apparently carried to these depths on a small lithospheric plate that subducted beneath the North European plate in Palaeozoic time². Subduction stopped and the sediments were forced back to the surface as the Austro-Alpine plate collided with the North European plate during the Variscan orogeny about 350 million years ago.

Although there has been other mineralogical evidence for subduction of sediment to mantle depths (for example, ref. 3), most of the evidence has been less direct. Coats⁴ proposed in 1962 that andesitic volcanoes associated with orogenic belts such as the Andes resulted from sediments being dragged down Benioff, or subduction, zones to mantle depths and melted. We now realize that the explosive nature of these types of volcanoes (Mount St Helens and Mount Pinatubo are examples) is caused by their high water content, some of which is derived from subducted oceanic crust and sediment. In retrospect though, Coats's idea was remarkably prescient, predating the concept of subduction implicit in plate-tectonic theory.

The first evidence in support of Coats's theory was the similarity of lead isotope ratios between these 'island-arc magmas' and oceanic sediments⁵. More dramatic confirmation of sediment subduction came later from beryllium isotope studies⁶. ¹⁰Be, an unstable isotope that decays with a half-life of 1.6 million years, exists on the planet only because it is continually created by interaction of cosmic rays with the atmosphere. The ¹⁰Be thus created is removed by rain and accumulates in sediments. That some island-arc magmas contain ¹⁰Be, albeit in minute quantities, whereas other magmas never do, demonstrates that sediment is subducted to depths of 100 km or more. Many other isotope studies have confirmed subducted sediment is involved in island-arc magma genesis, but they have also generally found that

sediments constitute no more than a per cent or two of the material that melts to form these magmas; the remainder of the magma source appears to be mantle peridotite.

Armstrong⁷ recognized quite early that recycling of sediment into the mantle could have profoundly influenced the evolution of both the crust and mantle. He argued that although volcanism has added material to the continental crust throughout Earth's history, these additions have been balanced by loss of crustal mass through subduction of sediment. The result, he claimed, is that the continental mass has remained essentially constant for the past 4,000 million years. Though the idea of constant crustal mass has not found wide acceptance, most geologists agree that significant volumes of continental crust have indeed been destroyed through subduction.

Mantle plumes, which rise from the deep mantle and partially melt to produce oceanic-island volcanoes such as Réunion, are chemically distinct from the remainder of the mantle. Hofmann and I proposed⁸ that this chemical distinction arises from recycled oceanic crust and sediment being present in plumes. We also speculated that higher concentrations of the radioactive, heat-producing elements potassium, uranium and thorium in this recycled material may be the cause of plumes. If, as is now commonly supposed, plumes originate at the core-mantle boundary, conductive heat transport across this boundary may be the dominant cause of plumes. Nevertheless, the idea that plumes contain deeply recycled crustal material has found wide acceptance.

The rocks Becker and Altherr studied were limestones, a reminder that mantle recycling of sediment may be important in another aspect of planetary evolution: climate. Atmospheric CO₂ may play an important, if not dominant, part in controlling Earth's climate. CO₂ is continually cycled through a number of reservoirs on the surface of the Earth. The atmosphere is among the smallest of these and carbonate rocks are by far the largest. Atmospheric CO₂ concentration depends largely on the fluxes between these reservoirs⁹, but it is also influenced by net losses or gains to the total surficial carbon system. Volcanism adds CO₂ to this system, and subduction of carbonate rock to the deep mantle removes it. CO₂ added by the extensive Cretaceous volcanism (due, perhaps, to the initiation of a number of new mantle plumes¹⁰) may have induced the climatic warming of that time. Since then, the

1. Alderson, K. L. & Evans, K. E. *Polymer* **33**, 4435-4438 (1992).
2. Milton, G. J. *Mech. Phys. Solids* **40**, 1105-1137 (1992).
3. Fung, Y. C. *Foundations of Solid Mechanics*, 353 (Prentice-Hall, Englewood, 1968).
4. Kolpakov, A. G. *Prikl. Mat. Mekh* **59**, 969-977 (1985).
5. Gibson, L. J., Ashby, M. F., Schajer, G. S. & Robertson, C. I. *Proc. R. Soc. A* **382**, 25-42 (1982).
6. Caddock, B. D. & Evans, K. E. *J. Phys. D* **22**, 1877-1882 (1989).
7. Lakes, R. S. *Science* **235**, 1038-1040 (1987).
8. Lakes, R. S. *J. Mater. Sci.* **26**, 2287-2292 (1991).
9. Evans, K. E., Nkansah, M. A., Hutchinson, I. J. & Rogers, S. C. *Nature* **353**, 124 (1991).
10. Cherkashev, A. & Gibiansky, L. *J. Mech. Phys. Solids* (in the press).

evolution of planktonic foraminifera may be shifting the net flux of CO₂ towards the mantle. The carbonate shells of these organisms cover much of the ocean floor, whence subduction into the mantle is a likely fate.

The Dunkelsteiner marbles are only the latest in a long trail of evidence of deep sediment recycling. They are nonetheless more dramatic than much of the other evidence, which has been based largely on extremely subtle variations in the abundances of trace elements and isotopes. More importantly, these rocks may provide an opportunity to study at first hand the reactions that subducting sediments undergo and thus solve a number of stubborn problems. The processes by which subducted sediments find their way into island-arc magma is still not understood. Do the sediments themselves melt, or do they dehydrate, with fluids carrying the sedimentary chemical signature into the overlying mantle peridotite undergoing melting? What fraction of subducted sediment is quickly returned to the surface through island-arc volcanism rather than being carried to the deeper mantle? Do some elements preferentially take

this shorter route while others take the deeper route?

The key to answering these questions is understanding the processes and reactions that occur as the subducting lithospheric plate passes beneath the island-arc volcanoes. What makes the Dunkelsteiner Wald marbles particularly interesting is that the maximum metamorphic pressures correspond to about 100–120 km, the typical depth at which the subducting plate passes beneath island-arc volcanoes. □

William M. White is in the Department of Geological Sciences, Cornell University, Ithaca, New York 14850, USA.

1. Becker, H. & Altherr, R. *Nature* **358**, 745–748 (1992).
2. Carswell, D. A. *Eur. J. Miner.* **3**, 323–342 (1991).
3. Smith, D. C. *Nature* **310**, 641–644 (1984).
4. Coats, R. R. in *The Crust of the Pacific Basin, AGU Monograph 6* (eds MacDonald, G. A. & Kuno, H.) 92–109 (Am. geophys. Un., Washington, 1962).
5. Armstrong, R. L. *Earth planet. Sci. Lett.* **12**, 137–142 (1971).
6. Tera, F. *et al. Geochim. cosmochim. Acta* **50**, 535–550 (1986).
7. Armstrong, R. L. *Rev. Geophys.* **6**, 175–199 (1968).
8. Hofmann, A. W. & White, W. M. *Earth planet. Sci. Lett.* **57**, 421–436 (1982).
9. Berner, R. A., Lasaga, A. C. & Garrels, R. M. *Am. J. Sci.* **283**, 641–683 (1983).
10. Larson, R. L. *Geology* **19**, 547–550 (1991).

QUANTUM OPTICS

Crystals with a light touch

Andrew Steane and Christopher Foot

THE recent remarkable progress in laser manipulation of atoms shows no signs of slowing. Two elegant experiments now give clear evidence of quantized motion of atoms confined in delicate potential wells set up by the laser light. The atoms move in a periodic potential created by a laser standing wave so that the motion resembles that of electrons in a crystal. In a crystal, the lattice is made of ions and produces a periodic potential for the motion of electrons; in the new experiments the lattice is formed by the light (standing wave), and quantum motion of whole atoms is obtained. The two experiments complement each other in that they use different atoms, and different methods to probe the atomic motion: the group at the Ecole Normale Supérieure in Paris (P. Verkerk *et al. Phys. Rev. Lett.* **68**, 3861–3864; 1992) measured the absorption spectrum of a weak laser beam propagating through a collection of caesium atoms, while at National Institute of Standards and Technology, Gaithersburg, P. S. Jessen *et al. (Phys. Rev. Lett.* **69**, 49–52; 1992) used an optical heterodyne method to analyse the fluorescent light from rubidium atoms.

We usually think of the motion of an atom as a classical variable. This works well for high-lying quantum states be-

cause of the correspondence principle, but understanding the new experiments requires a quantum mechanical picture of the motion. The important new factor in these experiments, in addition to the atoms' low energies, is that each atom is confined in a very small region, so that their motion is quantized, with the separation between different states of motion rising as $1/r^2$ as the radius of confinement, r , decreases. The optical standing wave formed by two counter-propagating laser beams of suitable polarization gives the atoms a lattice of wavelength-sized potential-energy wells.

The well depth depends on the light's intensity; the more intense the light, the deeper the wells. But the radiation also excites the atomic motion to higher energies. These two effects are balanced in the new experiments, which use diffuse laser beams of a few milliwatts power to give wells deep enough to contain a few bound vibrational levels.

The second requirement is to reduce the atoms' energy enough for them to fall into the quantum well. (Indeed, when the prospect of trapping atoms in the optical potential wells of a standing wave was first considered, it was concluded that the atoms would rapidly 'boil out' of the wells: quantum fluctuations in

Poles together

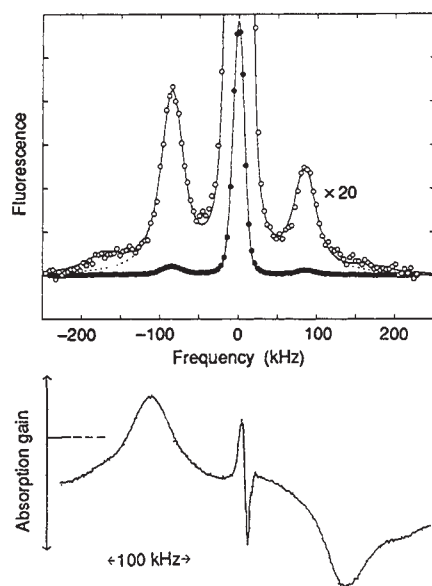
OPPOSITE electric charges attract and so would magnetic monopoles — isolated north and south poles — if they existed and were point-like objects. Magnetic monopoles are in fact predicted by many theories in elementary particle physics, but are extended, structured objects. As Y. Aharonov *et al.* note in the *Physical Review* (**D46**, 1877–1878; 1992) these monopoles certainly attract at a distance, but whether the force between them remains attractive at short distances is far from obvious. By finding a simple expression for the energy of two monopoles of opposite magnetic charge, the authors show that the energy always decreases with decreasing separation — there is always attraction. Curiously, though, the converse does not follow: extended monopoles of like charge may or may not maintain a repulsive force down to small distances, depending on their precise internal geometry.

Spectral zipper

THE spectrin family of structural, probably elastomeric proteins are made up of two antiparallel chains, which may be identical (α -actinin, dystrophin) or different (spectrin, fodrin), and consist of repeating units, typically of 106 residues. Whereas electron microscopy seemed to suggest that the two chains are united only at their ends, it is now clear that they interact along their lengths. D. W. Speicher *et al. (J. biol. Chem.* **267**, 14775–14782; 1992) now neatly establish that in red-cell spectrin the association depends on the first few (probably four) repeats at one end; these interact strongly and the remainder are then in register to close like a zipper, just as a DNA duplex forms by propagation of base pairing from an initially formed nucleus — a new kind of structural complementarity in proteins.

Nano-battery

NOT a year has passed since a micrometre-scale electric battery was devised (see *Résumé*, 31 October 1991), and already that record has been overturned. W. Li *et al. (J. phys. Chem.* **96**, 6529–6532; 1992) describe a nanometre-scale cell put together on graphite using a scanning tunnelling microscope. First, a pair of silver pillars just 150–200 Å across were deposited electrochemically beneath the microscope's two tips, from a solution of silver fluoride. Then (without disturbing the tips), the solution was changed for copper sulphate and copper pillars laid down nearby. Current flowed spontaneously through the graphite, dissolving the copper electrode and coating the silver 'cathode' with copper. For 45 minutes, the current remained at scarcely 5×10^{-18} A, displacing around 75,000 copper ions in all.



Crystal clear: two spectra reveal quantized motion of atoms trapped in an array of light-force-induced potential wells in a standing wave of laser light. Top, Jessen *et al.* recorded the spectrum of the trapping light scattered by the trapped atoms. Below, Verkerk *et al.* used a second laser beam to probe the atoms and found absorption at higher energies and stimulated emission at lower energies (broad dip and peak). In addition they observed a very narrow central feature which indicates long-range order over many wavelengths.

the light-atom interaction would raise the average energy above the well depth in a few microseconds.) This is also done with light, using laser cooling in which many photons scatter off a moving atom, slowing it down. This method reduces the temperature for the atomic motion to just a few microkelvin, low enough for most of the atoms to be in the lowest level of the well.

In each of the new experiments, a one-dimensional laser standing wave was used for simplicity. A cloud a few millimetres across, with 10^8 atoms cooled to a temperature of a few microkelvin, was placed in the standing wave. Jessen *et al.* probed the atoms by recording the spectrum of the scattered light (top figure). High resolution was obtained by mixing the atomic fluorescence with light at the same frequency as the incident laser beams. This optical heterodyne technique enables the small difference introduced in the scattering process to be seen directly on a Fourier transform spectrum analyser.

The central peak in the spectrum is light scattered at the laser frequency. Its width is less than the natural atomic linewidth; this alone implies that the atomic motion is localized to less than an optical wavelength (which causes narrowing of the spectrum by the Lamb-Dicke effect). The two sidebands are shifted by a frequency equal to the level

separation in the optical potential wells, and will be recognized by spectroscopists as Raman transitions which transfer the atom from one level to another in the well. A classical oscillator would also have sidebands, but the quantum nature of the motion here is shown by the asymmetry between the red and blue sidebands; the difference in strength is due to the population difference between adjacent levels. At the lowest temperatures achieved, the results imply that 60 per cent of the atomic population is in the lowest level of the well.

Verkerk *et al.* recorded the absorption spectrum of a second, weak beam directed almost parallel to one of the two travelling waves of the radiation field to identify their 'crystal' (lower figure). The frequency of the probe is scanned a few hundred kilohertz above and below the frequency of the standing wave (pump frequency). Once again there are two features in the spectrum which correspond to Raman transitions transferring an atom between levels in the quantum well. And they also indicate the presence of a significant difference in population between levels. When the probe frequency is 100 kHz above that of the pump, there is absorption of the probe through Raman transitions which transfer atoms from the lowest to the next higher levels in the wells. 100 kHz below the pump frequency, the probe is not absorbed but instead is amplified: it stimulates atoms excited to a higher level (by the standing-wave radiation) to radiate their energy and drop down again. Indeed, laser action (light amplification by stimulated emission) has been obtained off a similar transition using cold atoms in a magneto-optical trap by another group in Paris.

Between the two Raman features, there is an extremely narrow central structure. The explanation of this is more involved, but a crucial feature is that many wells of the periodic potential are involved, each having an atomic population. The pump beams are scattered through Bragg diffraction off the whole atomic distribution, producing a wave which interferes with the probe beam by two-wave mixing — the authors were observing the special quantum-optical properties of the collection of atoms, an atomic sample which is in a state intermediate between vaporous (or fluid) and crystalline. The magnetic properties of the sample are involved. For example, one can deduce that atoms in alternate quantum wells have their angular momentum aligned in alternate directions — the light-trapped lattice has an antiferromagnetic structure. □

Andrew Steane and Christopher Foot are in the Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, UK.

The wages of sin

THE mass media are widely feared. They can force publicity, unwelcome and even disastrous, on any individual, whose only defence is an expensive and uncertain libel action. More effective privacy laws seem impossible to formulate. Last week Daedalus proposed a computerized market in personal data, forcing bureaucrats to pay for their intrusive demands. He now plans to extend it to the media.

The principle is very simple.

Advertisers already pay the media a free-market price to publicize their claims. Conversely, people who are unwillingly publicized by the media should be legally entitled to a free-market payment.

Take, for example, that enduring favourite of the British tabloid press, the hapless vicar or religious dignitary caught in some sexual misdemeanour. Give the victim a legal right to bargain for recompense, and his claim is at once subject to market forces. Suppose the price is set too high. Vicars all over the country will leap into bed with their parishioners to claim the wages of sin. The market will be saturated, and the price will plummet. A going rate will soon emerge, at which the supply of naughty vicars exactly equals the demand. Thereafter, any vicar exposed by the press could automatically demand the going-rate payment. Rarely traded stocks, like those of amorous cabinet ministers, would of course be more volatile: even so, payment would be spread far more fairly than by the current system, in which a mass of disregarded victims is balanced by the occasional libel millionaire.

Privacy will thus acquire a fair market price. The media will remain uncensored, but will pay for what they now steal. Furthermore, payments will go to the victims, not the unsavoury crew of snoopers, informers and photographers who now shop them to the media for money. A criminal who shops his own accomplice, however, or a mistress who informs on her own lover, could be sued by the other party for a defined share of the takings.

A distributed computer-net, like those of the financial world, could easily convert thousands of such daily bargains into a set of going rates for the various categories of news. They will be most revealing. A simple photograph or interview, for example, might almost be free: people will be torn between embarrassment and pride. Exhibitionists may even pay to confess a night of shame, or boast of a crime. Daedalus will be interested to discover, for example, whether in strict cash terms a TV crew is more anxious to film a mob of rioters than the rioters themselves are to get on TV.

David Jones

Yeast retrotransposon revealed

SIR — Whereas considerable emphasis is being placed on the genes identified by large-scale sequencing projects, much will also be learned about genome organization and evolution. In particular, the nucleotide sequences of regions between genes will probably identify a wealth of transposable element insertions and increase our understanding of mobile genetic element diversity. An example of this has recently been provided by the DNA sequence of yeast chromosome III.

Oliver *et al.*¹ reported an open reading frame (ORF) near the left telomere of chromosome III (YCL74W) showing significant amino-acid sequence similarity to *Arabidopsis thaliana* retrotransposon

observed for other yeast retrotransposon families and result from recombination between LTRs of complete elements⁴.

The Ty5-1 LTRs share about 92% nucleotide identity, suggesting that the Ty5-1 insertion is ancient. Neither Ty5-1 nor Ty5-2 is flanked by target site duplications typically generated on integration of transposable elements. The Ty5 target site duplications may have been eroded by mutation or lost through recombination with LTRs elsewhere in the genome.

The organization of the internal domain of Ty5-1 indicates that this element has accumulated mutations that obscure its identity as a retrotransposon (see figure). Three main ORFs lie between the LTRs, including

4% of the yeast chromosome III is composed of transposable element sequences, and they may constitute an even larger fraction of the genomes of higher organisms.

Daniel F. Voytas

Jef D. Boeke

Department of Molecular Biology
and Genetics,
Johns Hopkins University School
of Medicine,
Baltimore,
Maryland 21205, USA

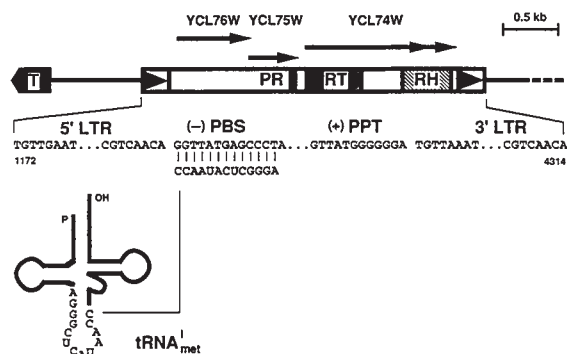
1. Oliver, S. G. *et al.* *Nature* **357**, 38–46 (1992).
2. Doolittle, R. F., Feng, D. F., Johnson, M. S. & McClure, M. A. *Q. Rev. Biol.* **64**, 1–30 (1989).
3. Burton, L. L. & Astell, C. R. *Molec. cell. Biol.* **6**, 1352–1356 (1986).
4. Boeke, J. D. & Sandmeyer, S. B. in *The Molecular Biology of the Yeast Saccharomyces* (eds Broach, J. R., Pringle, J. R. & Jones, E. W.) 193–262 (Cold Spring Harbor, New York, 1991).
5. Kikuchi, Y., Yumiko, A. & Shiba, T. *Nature* **323**, 824–826 (1986).

Alternative blood substitutes

SIR — Otto *et al.*¹ draw attention to the possible problem of increased susceptibility to bacterial infection in patients treated with haemoglobin-based blood substitutes. Their main point is that many bacterial pathogens can use extracellular haemoglobin as an iron source, and this may potentiate infection in the recipients of such materials.

Although haemoglobin is an excellent and obvious choice as the basis of an oxygen-transport system for intravascular use, it is not the only approach. The main alternatives are perfluorochemicals (PFCs), compounds which are chemically inert and have the capacity to dissolve large volumes of respiratory gases^{2,3}. The carbon-fluorine bond in PFCs is very strong (about 116 kcal mol⁻¹) and this confers exceptional chemical and thermal stability, and hence the problems discussed by Otto *et al.* are unlikely to arise *in vivo*.

Interest in the use of PFCs as biological carriers of oxygen arose when it was shown that small mammals could survive by 'breathing' an oxygenated PFC liquid in which they were immersed⁴. One problem with PFCs is that they are water-immiscible and so must be emulsified before intravascular use. Although emulsions of PFCs were first tested on humans more than 10 years ago, they have not become widely used. One reason is that the first-generation formulations, which typically contained only 10% by volume of PFCs, had relatively low oxygen-carrying capacities (about 6% of whole blood at ambient oxygen tension), and consequently demanded concurrent oxygen therapy. Acute adverse reactions in the first hu-



Structural organization of the telomeric Ty5-1 insertion. T, chromosome III telomeric repeats (5' C₁₋₃A 3'). Boxed triangles, Ty5-1 LTRs, with LTR end-sequences shown below the cartoon. Arrows above the internal domain of Ty5-1, ORFs described in ref. 1; arrowheads represent stop codons; PR, protease; RT, reverse transcriptase; RH, RNase H. Sequences are shown for the putative Ty5 priming sites and complementarity to the anticodon loop of yeast initiator methionine tRNA; PPT, polypurine tract. Sequence coordinates are from ref. 1.

proteins and proteins from other *copia*/Ty1 group elements². Further analysis indicates that this ORF identified a previously unreported yeast retrotransposon that we designate Ty5.

The telomeric Ty5 insertion (Ty5-1) resembles other virus-like retrotransposons and retroviral proviruses (see figure). It is flanked by long terminal direct repeats (LTRs) ending in short inverted repeats and shares LTR end-sequences with all previously characterized retroviral LTRs (5' TG...CA 3'). In a previous study of the chromosome III left-arm telomere³, the sequence for the Ty5-1 5' LTR was reported and considered part of the type X telomere-associated repeat. Subsequently, Oliver *et al.*¹ designated the Ty5-1 3' LTR as a 'pseudo-X' repeat (residues 4,065–4,314). We also note that the other 'pseudo-X' repeat identified on chromosome III (residues 290,656–290,901)¹ is actually a solo Ty5 LTR (Ty5-2). Solo LTRs are frequently

observed for other yeast retrotransposon families and result from recombination between LTRs of complete elements⁴.

An interesting feature of Ty5-1 is its putative priming site for minus-strand DNA synthesis by reverse transcriptase. Retrotransposons and retroviruses prime minus-strand DNA synthesis from the 3' OH of a cellular transfer RNA that hybridizes to a short stretch of complementary nucleotides immediately adjacent to the 5' LTR (the primer binding site, PBS). The *copia* elements of *Drosophila* use a novel primer for minus-strand synthesis — an initiator methionine tRNA fragment⁵. The Ty5-1 PBS begins at the same position in the anticodon loop as the *copia* primer and is complementary to 13 base pairs of the yeast initiator methionine tRNA. The striking similarity between the PBS of these retrotransposons from such disparate organisms suggests this unusual priming mechanism may be widespread.

Including the Ty5 elements, more than

man trials^{2,3} inevitably reinforced doubts that PFCs would be useful or even desirable.

There have been two main developments in this area during the past 5 years. The first is improvement of the oxygen-carrying capacity, by the development of second-generation, high-phase-volume (greater than 50%) emulsions with excellent stability at room temperature and reduced toxic effects due to the use of purified components⁵. Second, PFCs and their emulsions have been used in a broad range of applications, for example, organ perfusion and preservation, diagnosis, cancer therapy, ophthalmology and cell cultures^{2,3}.

We propose a further therapeutic use for PFC emulsions, as controlled inducers of metabolic enzymes. Many substances induce hepatic enzymes (for example, phenytoin and some barbiturates), but they do not usually have therapeutic value owing to their other pharmacological side-effects. Particulate material, such as colloidal carbon, which can similarly induce metabolic enzymes following uptake into the liver, is retained in the body for long periods because it cannot be metabolized and thus is not used therapeutically.

Some PFCs can induce hepatic enzymes⁶ without these side-effects. PFCs appear to have little or no pharmacological action and they are excreted (as vapour via the lungs) over a period of days following injection at a rate which depends primarily on the molecular mass of individual compounds^{2,3}. The degree of enzyme induction by PFCs is related to their physico-chemical properties⁷, and it should therefore be possible to use different PFC emulsions selectively to induce specific enzyme pathways. This could be of considerable therapeutic benefit, for example, in the treatment of paracetamol poisoning, where the production of toxic metabolites (produced via the mixed function oxidase system) could be eliminated if the glucuronidation pathway could be activated to a greater extent than the mixed function oxidase system.

Kenneth C. Lowe

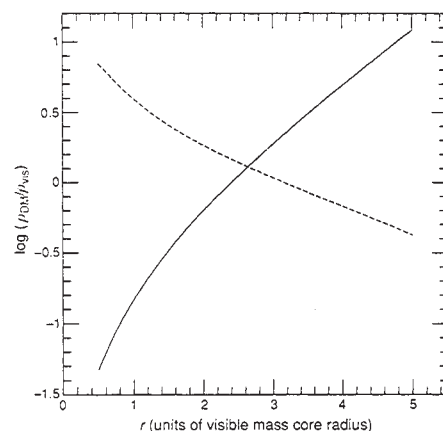
Clive Washington*

Departments of Life Science
and *Pharmaceutical Sciences,
University of Nottingham,
University Park,
Nottingham NG7 2RD, UK

Decaying neutrino theory

SIR — It is widely believed that Sciama's¹ decaying neutrino theory has been disproved by the failure of David-*et al.*² to detect an unidentified emission line at ~ 15 eV from the dark matter in the cluster of galaxies A665. But recent X-ray data³ for A665 favour models in which the dark matter in the regions of the cluster observed by David-*et al.* is mainly baryonic. So the decaying neutrino theory continues to be viable.

The X-ray data for A665 have been obtained by Hughes and Tanaka³ using



The local (diffuse dark)-to-visible density ratio as a function of a radius (in units of the scale-size, λ , of the visible mass distribution) between $\frac{1}{2}\lambda$ and 5λ for the spiral galaxy NGC 3198 (solid line; ref. 5) and for the cluster of galaxies A665 (dashed line; ref. 3). The scale-size λ is equal to the disk exponential length for NGC 3198 and to 1.3 times the X-ray core radius for A665.

the Ginga satellite. Combined X-ray and optical data have allowed them to construct a detailed self-consistent mass model of A665. This model includes: (1) a galaxy component of radial profile given by $\rho_G \sim (1+(r/530 \text{ kpc})^2)^{-1}$; and (2) an X-ray emitting, hot gas component of radial profile $\rho_X \sim (1+(r/380 \text{ kpc})^2)^{-1}$. Their preferred resulting gravitational mass distribution, $\rho_b \sim (1+(r/298 \text{ kpc})^2)^{-1.36}$, implies the presence of a dark-matter component more concentrated than the two visible ones. From their mass model we compute a quantity γ , defined as the ratio between the mass in diffuse dark matter and the mass in gas and in galaxies (inclusive of their dark haloes). We then find that γ goes from ~ 11 at $r=0.2$ Mpc to ~ 2 at $r=2$ Mpc. Thus in A665 the diffuse dark matter is more concentrated than the visible matter by an order of magnitude. More conclusively, the relative distribution of dark and luminous matter in clusters is opposite to what is found in galaxies⁴. In the well-studied exemplary

case of the spiral galaxy NGC 3198, the dark-to-luminous mass ratio increases markedly going outwards in radius⁵. To highlight this dramatic difference, in the figure we plot the local dark-to-visible density ratio between one-half and five times the typical size of the visible matter distribution for A665 (dashed lines) and NGC 3198 (solid line), using the mass model solutions of refs 3 (already described) and 5, respectively. At the edge of the optical structure the dark-to-visible density ratio in galaxies is ~ 10 times bigger than in clusters, whereas at the centre it is ~ 200 times smaller than in clusters. This supports the view that the nature of dark matter is different in clusters and in galaxies.

A similar high concentration of dark matter at the centres of clusters has been derived for the Perseus⁶, Coma⁷ and A85 (ref. 8) and A2199 (ref. 8) clusters and from gravitationally lensed arcs in A370 (ref. 9) and other clusters. It has been argued^{6,7} that this concentration probably implies that the dark matter in these clusters is dissipative, and therefore baryonic.

We conclude that David-*et al.*'s failure to detect an emission line from decaying neutrinos in the core of A665 cannot disprove in a straightforward way the decaying neutrino theory, in that the existence of substantial quantities of diffuse nonbaryonic dark matter in clusters has still to be confirmed.

Dennis W. Sciama*

Paolo Salucci

International School for Advanced Studies,
and *International Centre
for Theoretical Physics,
Strada Costiera 11, 34014 Trieste, Italy
Massimo Persic
Osservatorio Astronomico di Trieste,
Via Tiepolo 11, 34131 Trieste, Italy

- Sciama, D. W. *Astrophys. J.* **364**, 549–554 (1990).
- David-*et al.* *Nature* **351**, 128–130 (1991).
- Hughes, J. P. & Tanaka, Y. *Astrophys. J.* **398**, 62–68 (1992).
- Persic, M. & Salucci, P. *Astrophys. J.* **355**, 44–51 (1990).
- van Albada, T. S., Bahcall, J. N., Begeman, K. & Sancisi, R. *Astrophys. J.* **295**, 305–313 (1985).
- Eyles, D. *et al.* *Astrophys. J.* **376**, 23–32 (1991).
- Briel, U. G., Henry, J. P. & Böhringer, H. *Astr. Astrophys.* **259**, L31–L34 (1992).
- Gerbal, D., Durret, F., Lima-Neto, G. & Lachièze-Rey, M. *Astr. Astrophys.* **253**, 77–88 (1992).
- Bergmann, A. G., Petrosian, V. & Lynds, R. *Astrophys. J.* **350**, 23–35 (1990).

Scientific Correspondence

Scientific Correspondence is a relatively informal section of *Nature* in which matters of general scientific interest, not necessarily those arising from papers appearing in *Nature*, are published. Because there is space to print only a small proportion of the letters received, priority is usually given according to general interest and topicality, to contributions of fewer than 500 words, and to contributions using simple language.

- Otto, B. R., Verweij-van Vught, A. M. J. J. & MacLaren, D. M. *Nature* **358**, 23–24 (1992).
- Lowe, K. C. *Vox Sang.* **60**, 129–140 (1991).
- Lowe, K. C. *Clin. Hemorheol.* **12**, 141–156 (1992).
- Clark, L. C. & Gollan, F. *Science* **152**, 1755–1756 (1966).
- Bentley, P. K. *et al.* *J. Pharm. Pharmacol.* **41**, 661–663 (1989).
- Ravis, W. R., Hoke, J. F. & Parsons, D. L. *Drug Metab. Rev.* **23**, 375–411 (1991).
- Lowe, K. C. & Armstrong, F. H. *Biomater. Art. Cells Immob. Biotech.* (in the press).

The end of geohistory?

Gordon L. Herries Davies

From Stone to Star: A View of Modern Geology. By Claude Allègre. Translated by Deborah K. Van Dam. *Harvard University Press*: 1992. Pp. 300. \$39.95, £31.95.

I LIE upon the psychiatrist's couch. I listen as my present character is explained in terms of my childhood experiences. So it is, too, with our Earth. It lies there impassively while geologists clamber around like those Lilliputians who explored the recumbent Gulliver. And then all is explained — explained in terms of the geohistorical events of the telluric past. Of those events we can have no direct experiential knowledge. The past, like yesterday's dawn, is gone, never to be reclaimed. It is there to tantalize us like the crock of gold at the end of the rainbow. What lies within the grasp of the geohistorian is nothing more substantial than human beliefs about the past. Those beliefs, let it be noted, grow not out of the crust of the Earth itself, but rather out of the cerebra of that crust's Lilliputian observers.

Because geohistory originates within human minds, it is hardly surprising that the accepted styles of geohistory — like the accepted styles of architecture — have varied through the ages. This point is nicely illustrated by the changing perception of the relationship between terrestrial and extraterrestrial events. During the seventeenth and eighteenth centuries, 'externalist geohistory' prevailed. Such postulated geohistorical events as the Deluge were commonly explained in terms of cometary encounter or divine intervention. During the nineteenth century, the rise of uniformitarianism engendered 'internalist geohistory' and all extraterrestrial causation was declared taboo. Even Arizona's Barringer Crater was explained in terms of volcanism rather than meteorite impact. Finally, today, we see a neocatastrophism that looks heavenward in search of external causation for all manner of geohistorical events, from marine transgressions to the extinction of the dinosaurs.

Those who seek for relationships between science and society perhaps here have grist for their mill. When we feel puny and threatened — when we feel helpless in the presence of the Plague, the Lisbon earthquake of 1755, AIDS or the greenhouse effect — do we reflect our insecurity by resorting to externalist geohistories in which we are the microcosmic victims of cosmic forces? When we feel supremely self-confident — as did so many of Queen Victoria's contemporaries — do we enthrone the sub-lunar world as the seat of all that is geohistorically significant?

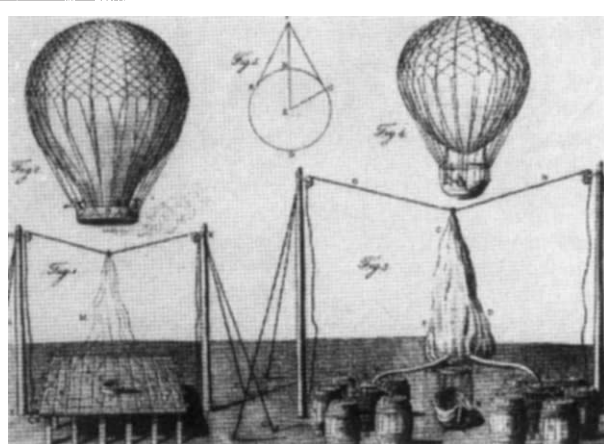
Professor Allègre is a distinguished French geochemist and an ardent disciple of the externalist school of geohistory. In this book, he protests at the myopia that has caused geologists to rivet their attention solely upon the events of the past 600 million years. He jettisons the tools of such a miniaturist approach and mounts a scaffolding from which he applies broad brush strokes to a theatrical set of massive proportion. His drama takes us out from Antarctic meteorites and the lunar mare, on to the canyons of Mars and the helium rains of Saturn. He shows how within our terrestrial laboratories the evidence derived from our space programmes may be processed to illuminate geohistorical problems that conventional geology could only dismiss as beyond solution. In short, Allègre seeks to bridge the gulf between astronomy and the Earth sciences. If you are a traditional geologist for whom plate tectonics represents the ultimate thrill, then take your stance upon Allègre's bridge and grow giddy at the prospect revealed before you.

The author — and his translator — are to be congratulated on the production of a book that is both highly readable and deeply informative. It is an excellent introduction to the exciting geohistorical discoveries made in recent decades and it will serve alike the needs of the geologist and the general reader of science. The book should, nevertheless,

carry the following health warning: 'Readers of this volume who are historians of the geosciences will find that this book may induce severe palpitations.' Its earlier chapters are riddled with historical misrepresentations and factual errors. A few specimens of these regrettable blemishes will serve to alert the reader to be *en faction*. Robert Jameson's name is misspelled (page 7). Kirwan did not discover the 'fossiliferous' Portrush sill (page 8). Steno's *Prodromus* does not date from 1671 (page 12). The Anglican church did not have "a great deal invested in geology" (page 13). Henry Cavendish was never ennobled (page 19). There is confusion over William Smith's geological cartography (page 40). It is untrue to say that Rutherford emigrated from New Zealand to Canada and thence to England (page 46). Such flaws are irritating, but their presence I forgive because I am satisfied that the volume is a useful addition to our literature.

As I read the book I was repeatedly reminded of Thomas Burnet's *Telluris Theoria Sacra*, first published in 1681. Both books ambitiously trace the history of our Earth in its cosmic setting and even some of Allègre's diagrams carry strange echoes of Burnet's illustrations. In its day, Burnet's theory earned wide currency (even Newton was impressed), yet today it lies rusting upon the scrapheap of science. In Allègre's work I detect a current of the same triumphalism that surrounded Burnet. Allègre seems to feel that the remarkable and mutually supportive discoveries made in the past few decades have brought us close to the end of our geohistorical journey. It has been reserved for us to attain the peak of that mountain where previous generations have merely strug-

EARLY hot-air (left) and hydrogen (right) balloons (1785). It was Henry Cavendish's estimate of the density of hydrogen in 1766 that led Joseph Black to suggest filling balloons with the gas, which had been made cheap by John Roebuck's commercial production of sulphuric acid. The first trial hydrogen balloon was released in Paris in 1783, and interest in balloons soon became enormous. King George III offered to subsidize 'aerostatic experiments' if they were undertaken by the Royal Society, whose rash reply was that "no good whatsoever" could come of them. Yet within years, balloons were widely used for scientific purposes, whether to calculate the "powers of ascension" or to explore "the upper regions". In 1784, the physicist John Jeffries used a balloon to collect samples of upper air for Cavendish to analyse, and James Watt and Matthew Boulton later used one to investigate the cause of the growling of thunder. Not surprisingly, within months of the balloon's invention, people had begun to suggest its use in warfare. A chapter on balloons appears in *The Chemical Revolution: A Contribution to Social History* by A. Clow and N. L. Clow, a work originally published in 1952 and now reprinted in paperback by Gordon and Breach. \$28, £15.



gled blindly upon the flanks. But is such an interpretation really justified? From Burnet onwards every succeeding generation has felt just as confident in its geohistorical beliefs as we feel in ours. Their confidence was misplaced. Is ours? We reject today the geohistories devised by all ages before our own. Why, then, should we expect future generations to cherish our geohistories as representing ultimate truths? Allègre's book is a useful summary of the present state of our geohistorical thinking. Time alone will reveal whether our modern thinking has a greater durability than that enshrined in Burnet's eloquent theory of 1681. □

Gordon L. Herries Davies is at Trinity College, University of Dublin, Dublin 2, Ireland.

Chocolate kookies

Peter A. Lawrence

Models for Embryonic Periodicity. By Lewis I. Held Jr. Karger: 1992. Pp. 119. SFr. 195, DM 234, £84.80, \$156.

THE long-lived rift between genetics and embryology stems from the different natures of the two subjects. Genetics is a hard discipline full of calculations, statistics and measurements; even morphology is transformed into numbers representing heritability and expressivity. Embryology, at least until recently, has bordered on the philosophical and has featured wonder as much as analysis. The two disciplines have always attracted different kinds of scientists and this contributed in the past to the mutual distaste that delayed the disciplines' inevitable fusion.

One characteristic of embryologists, particularly those interested in the fundamental problem of pattern formation, is their propensity to invent theoretical models that aim to explain the mechanisms behind such diverse matters as patterning of butterfly wings, the orderly connections between retina and brain or the stripes of a zebra. Generally, perhaps because of the whimsical nature of embryologists, the models are rather loosely set out and live explanations of them are often accompanied by hand-waving and kookie looks. In this original little book, Lewis Held, whose style in science is more that of a geneticist than of an embryologist, has brought his incisive intelligence and background in computer programming to bear on a whole menagerie of models. He has scoured the literature with evident enthusiasm; he takes each model without favour,

holds it up and shakes it until the bits and bobs fall off, and presents it to us naked and unadorned. He explains the difference between related models in simple language and adds helpful diagrams and tables. The result is a very useful guide to what, no doubt, is a great deal of nonsense with a few valuable ideas mixed in.

For example, what are the essential differences between the prepattern model of Stern and the gradient-based models of Sander, Wolpert and others? If you have struggled with this conundrum in front of a class of mutinous students you will know how valuable it would be to have a copy of Held's guide at hand. The polar coordinate model (designed to explain patterns of regeneration in insect and vertebrate limbs), the epigenetic landscape of Waddington, Sperry's chemoaffinity model, Kauffman's model of transdetermination and many many others are all here, inhabiting the book like animals in a zoo. Some are well adapted and successful and others barely cling to life. All are given even-handed treatment as if there were no higher criteria that could judge them or junk them. Held has an entertaining and simple style; he uses a pleasingly large vocabulary and manages to avoid the heavy jungle jargon that becomes more and more commonplace (take a sideways look at the titles of papers in *Proceedings of the National Academy of Sciences*, for instance). Sometimes he achieves an almost startling clarity that makes the original papers he discusses look unbearably fuzzy and even makes some of the models look a bit ridiculous. But this is no bad thing.

The subjects dealt with cover a great deal of ground and include many organisms from alligators, *Anabaena* and *Antirrhinum* to *Xenopus* and zebras. The text is extensively referenced, and the list of more than 1,000 papers on patterns at the end will prove extremely useful. It is a pity there is no index.

This book will be welcomed by researchers who have a particular interest in the biology of pattern formation. It is mainly *theoretical* biology, because only a few experiments are mentioned and it is these that can allow one to evaluate models or even destroy them. But, of course, there wasn't room for the experiments and adding them would have obscured Held's purpose. The result is a short expensive book that is easier to understand than to pay for. It is not really a full scientific meal but, like a fancy box of chocolates, it makes a welcome addition to one. □

Peter A. Lawrence is in the Laboratory of Molecular Biology, Medical Research Council, Hills Road, Cambridge CB2 2QH, UK.

To the heart of the matter

David Weatherall

Vital Circuits: On Pumps, Pipes, and the Workings of Circulatory Systems. By Steven Vogel. Oxford University Press: 1992. Pp. 315. £19.95, \$24.95.

IN one of the many asides in Steven Vogel's new book, there is an insert from the diary of Samuel Pepys, dated 21 January 1665. "Before I went to bed I sat up until 2 o'clock in my chamber reading Mr. Hooke's microscopical observations, the most ingenious book I have ever read in my life." Pepys' fascination with microscopy evokes feelings of nostalgia in Dr Vogel as he contemplates an age during which a newly published scientific treatise such as *Micrographia* was seized upon by the intellectual community at large. Sadly, he muses, we have progressed to the point at which the community is no longer interested, the treatise no longer intelligible, and when commercial bookstores would not dream of stocking such items. He is right, of course, which is at least one good reason why scientists who are suitably gifted should try to make their field accessible to a general audience. On the evidence of this book, Vogel has certainly succeeded in this task.

Vogel has set out to explain the complex workings of the heart and vascular system in a way that will be comprehensible to nonexperts. Using simple models derived from household plumbing, and examples culled from a wide variety of species, he leads us through the functions and regulation of the heart and circulation, and deals at length with the thorny problems of fluid mechanics and blood viscosity — topics which are so often ducked by teachers of physiology — with extraordinary clarity and lightness of touch. The sections on capillary circulation and oxygen transport are particularly well done. As part of our cardiological education, we are urged to dissect animal hearts for ourselves; a list of suppliers is included. We are even instructed on what to do with the hearts at the end of the dissection: there is an excellent recipe for a vindaloo that requires two pounds of diced ventricle. Every chapter is enlivened with a series of entertaining footnotes, the subject matter of which ranges from the reasons why Chaucer and Harvey use the female gender to describe nature to the dimensions of the *os penis* of the whale, a bone which, incidentally, seems to have changed its size quite remarkably in the 30 years since one was waved at me by an enthusiastic anatomist

in my medical-student days.

It is difficult to find fault with this eccentric and lively account of the circulation. If I have any quarrel at all with Vogel it is in his apparent lack of regard for the red blood cell which, he suggests, is not a proper cell because it does not have a nucleus. But surely a cell that travels many hundreds of miles in conditions of unbelievable turbulence, is dragged through capillaries half its dimension and then crammed together with its fellows in the osmotic backwaters of the spleen, maintains itself and its haemoglobin against all the odds in this unfriendly environment yet is able to adapt its function to prevailing oxygen requirements, and suffers the final indignity of being eaten by a macrophage, is worthy of a little more respect (and space). But this is a personal bias.

Like all good storytellers, Vogel leaves a few intriguing questions hanging in the air. How, for example, does he know that the systolic blood pressure of an energetic octopus is about 80 mmHg, and who is it (or perhaps was it) that discovered why arboreal snakes do not faint when they climb trees?

The true mark of a really good popular science book is that it is as interesting to the expert as it is to the general audience to whom it is addressed. I strongly recommend this book to anybody who teaches the physiology of the circulatory system, and to cardiologists and all those who live in fear of them. □

David Weatherall is at the Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK.

Understanding historical climate

William F. Ruddiman

Climate Since AD 1500. Edited by Raymond S. Bradley and Philip D. Jones. Routledge: 1992. Pp. 679. £85.

WHAT can the climate changes of the past 500 years tell us about the future? To address this important question, the authors of this volume have assembled many different records of past global climate change and reviewed the limitations on interpreting these data. The result is a most valuable compilation.

Three kinds of data are examined, all from the continents or their margins. Historical data include recent instrumental recordings and a wide variety of earlier observations. Coverage before 1880 is adequate for portions of the northern continents, but there are large gaps in the data for North Africa and Asia, and little information for the Southern Hemisphere. Dendroclimatological data come from studies of annual tree-ring width and variations in density for many trees at each site and borings within each tree. In recent years, there has been fairly wide geographical coverage in the continents of the Northern Hemisphere and pioneering investigations in the Southern Hemisphere. Climate can also be reconstructed from ice cores. Glacial ice varies greatly in its physical and chemical composition, and the chronologies developed from annual layering can be verified against historically recorded volcanic eruptions.

But several factors limit the use of these methods. Historical records are often semiquantitative and are subject to complex observer biases. Also, the diverse array of recorded parameters (rain in east Asia, sea ice in North America)

makes comparison difficult. Dendroclimatology studies are limited by, for example, the difficulty of allowing for long-term growth trends and, in some regions, having to calibrate and verify against only brief historical records of climate. For ice cores, the causes of variations are often not well understood (for example, the relative roles of air temperature and water-vapour source in changes in oxygen isotope ratios).

An underlying concern of the volume is the relevance of the changes in climate during the past 500 years to present and future climate, particularly those changes caused by greenhouse gases. If scientists could identify significant patterns of climate change over this period and link them unambiguously to known forcing functions, it should be possible to recognize some of the 'natural variability' in the climate system and to detect more accurately the magnitude of man-made changes, both in the past and future.

The baseline for calculating this century's global warming happens to fall in the relatively cool late 1800s, when instrumental recording became sufficiently broad for climatologists to attempt estimates of global temperature. Yet the many indicators of climate given in this volume point to substantially warmer temperatures in earlier times, such as the 1700s. If these warmer temperatures (or some average) were used as the baseline, the inferred warming in this century would be greatly reduced. This would aggravate the existing problem of explaining why warming this century is less than expected from observed increases in trace gases.

Answers to this broader question are not yet at hand. The challenge of synthesizing the data in this volume into a



Rush-hour traffic creeping across the Bay Bridge, San Francisco. Sixty per cent of the oil burned in the United States each day is consumed by cars and trucks, adding hundreds of thousands of tons of CO₂ to the global greenhouse. The picture is taken from *Global Warming: Understanding the Forecast* by Andrew Revkin, published by the American Museum of Natural History Environmental Defense Fund. \$29.95, £20.

coherent view of global or even regional climate change is daunting, in part because different indicators are sensitive to different climate indices and seasons (for example, summer precipitation and winter temperature). In addition, geographical coverage for the continents is still limited, and most ocean regions will never yield annual climate indicators. Climatic forcing functions are also not yet well understood, including solar variability (for which no unambiguous evidence exists of a climatic effect), volcanic eruptions (the climatic effects of which are here given surprisingly lukewarm support; a stronger case has been made elsewhere by Steven Porter) and El Niño oscillations.

Nevertheless, this volume succeeds admirably in bringing together many of the multifaceted data from which future progress can proceed. □

William F. Ruddiman is in the Department of Environmental Sciences, University of Virginia, Charlottesville, Virginia 22903, USA.

Bedside manners

Robert G. Petersdorf

The Crisis in Clinical Research: Overcoming Institutional Obstacles. By Edward H. Ahrens Jr. *Oxford University Press: 1992. Pp. 236. £27.50, \$29.95.*

WHAT is clinical research? What was termed clinical research when I was a student bears little semblance to what is termed clinical research today. Most work then consisted of questioning patients at their bedsides, examining samples collected from patients with disease, relating this to experimental animal models and doing clinical trials.

In this book, Edward Ahrens, an emeritus professor at The Rockefeller University and a clinical investigator for some 40 years, looks at the shift in emphasis in clinical research in the United States from patient-oriented research to research at the cellular and molecular level. In searching out the reasons for this change, he carefully examines the institutional supports for clinical research, including the medical schools and clinical centres in which the research is carried out and the main funding source, the National Institutes of Health (NIH). Ahrens argues that the traditional roles of medical schools — teaching, basic research and providing service to patients — have become so seriously out of balance that US medicine as a whole is in jeopardy. He concludes by prescribing several recommendations aimed at correcting these imbalances.

Research definitions

Unlike others, Ahrens permits no ambiguities in his definition of clinical research. He categorizes it into several kinds of activity, examining each in detail. These include the study of the mechanisms, modelling and management of human diseases, the development of new technologies and the assessment of health care. And he makes a compelling case for why this research is absolutely essential. Among his reasons are that there are no perfect animal models, particularly for chronic degenerative diseases and many infections; that new diseases will continue to appear (AIDS is a mordant reminder); that gene mapping and other reductionist approaches often eventually require the study of diseases in whole humans (cystic fibrosis is one example); and that doctors need to be able to make sound observations and to question why and how clinical phenomena occur.

Yet during the past 15 years there

have been progressively fewer grants awarded to investigators for patient-oriented research; and currently, only half these grants go to MD investigators. In explaining this shift from bed to bench, Ahrens points to the fact that medical students now spend less time studying biochemistry, pharmacology and physiology in the laboratory, and more on rote memorization than on problem solving. With mounting debts, they are enticed more readily into the high-paid careers of clinical practice than into ones as academic physicians. Other disincentives include the increasingly inadequate supervision of novice clinical investigators, particularly in large departments, and the fact that it is more difficult to have papers published on clinical than nonclinical research.

Ahrens believes that both the NIH and medical schools are responsible for throwing serious obstacles in the way of clinical research. The NIH has become a research colossus whose accountability has been called into question. Ahrens suggests that the institutes must establish an outside review mechanism over and above the institutional review groups currently used to assess extramural programmes. He provides some evidence that grant applications for clinical projects are judged by the institutes more on whether valid experimental data will be obtained, than on the originality and inventiveness of the proposals.

As the largest general clinical research centre in the world, the NIH's Warren Grant Magnuson Clinical Center in Bethesda, Maryland, should be the pacesetter for this kind of research. But there have been no internal systematic reviews of the scientific merit of bedside research at the centre. In 1985, though, the Clinical Center was reviewed by a committee chaired by Donald Seldin. The committee showed that "there was substantial variation in the quality of the [research] protocols reviewed from truly outstanding to quite poor — some proposals would not pass the scrutiny of the extramural peer review process."

Ahrens is critical of how the Clinical Center's budget is determined by the individual institutes. Because of the escalating costs of clinical research due to new technology and therapeutics, several directors of institutes have discouraged bedside research. As a result, bed space at the centre has decreased and bed costs have increased. Ahrens suggests that the centre should be allocated a separate budget, independent of those of the institutes. Bedside researchers from the institutes would then compete for funds not with their NIH colleagues who do purely laboratory research, but with each other, as do researchers in clinical centres of medical schools. The most interesting suggestion

made by Ahrens is that patient-oriented basic research should be judged by groups of investigators who are experienced in clinical research.

At medical schools, however, research is only one of the activities on the agenda: teaching and providing services to the public are also essential roles. Clinical departments in most US medical schools are so large as to preclude any collegiate atmosphere and decent supervision, and have all too often become factories for rendering service, but only at a price. There is less time and money for clinical investigators to do research, to teach and to train. And, as Ahrens shows, the criteria for tenure and promotion disadvantages clinical investigators.

Training reforms

Ahrens proposes several reforms that he believes would improve the quality of clinical training. On the basis of Malthusian theory, he shows that productive investigators clone themselves at least once yearly. He fears that well-trained scientists, with research careers lasting say 10–15 years, are in danger of outgrowing the ability of the government to fund them. He indicates ways of making research training more streamlined and productive by establishing working partnerships between clinically skilled MDs and technically trained PhDs.

Ahrens has done a superb job in explaining how the NIH works, how grants are handled and what the problems with the peer-review system are. He backs up commonly held assumptions with the results of studies both inside and outside the NIH. By using questionnaires and interviews and holding a series of two-day meetings with various authorities on clinical research, Ahrens has produced a compendium that is full of data, yet a lot more interesting than the usual NIH guides.

If the book has a weak point, it is Ahrens' attempts in the first two chapters to put clinical research in the context of US medicine. He does not really deal with US medicine in global terms; rather, his commentary covers US medical education. But even here he falls short. His arguments are often dated and he intermingles his databases somewhat indiscriminately.

Nevertheless, Ahrens' conclusion that there is a crisis in clinical research is inescapable. He has done the entire research community a favour with his acute analysis. Every clinical investigator, research administrator and medical educator should be required to read this splendid book. □

Robert G. Petersdorf is at the Association of American Medical Colleges, 2450 N. Street NW, Washington DC 200037-1126, USA.

NATURE · VOL 358 · 27 AUGUST 1992

Seasonal and interannual variations in atmospheric oxygen and implications for the global carbon cycle

Ralph F. Keeling & Stephen R. Shertz

National Center for Atmospheric Research, PO Box 3000, Boulder, Colorado 80307-3000, USA

Measurements of changes in atmospheric molecular oxygen using a new interferometric technique show that the O₂ content of air varies seasonally in both the Northern and Southern Hemispheres and is decreasing from year to year. The seasonal variations provide a new basis for estimating global rates of biological organic carbon production in the ocean, and the interannual decrease constrains estimates of the rate of anthropogenic CO₂ uptake by the oceans.

THE dominant processes influencing the abundance of atmospheric O₂ on timescales less than many thousands of years are chemical reactions involving the creation and destruction of organic matter: that is, photosynthesis, respiration and combustion. These reactions cause atmospheric O₂ to vary inversely with atmospheric CO₂ with an O₂:C exchange ratio that depends on the composition of the organic matter. Additional chemical reactions, however, influence the abundance of atmospheric CO₂. The difference in the geochemistry of O₂ and CO₂ arises because, whereas dissolved O₂ is chemically neutral, dissolved CO₂ reacts with water to form carbonic acid, which in turn reacts with basic compounds such as carbonates. As a result of the differing acid-base properties of O₂ and CO₂, measurements of changes in atmospheric O₂ can constrain aspects of the global carbon cycle that are not constrained by measurements of atmospheric CO₂ alone.

Consider, for example, the processes controlling the increase in atmospheric CO₂ over the past century. The dominant sources of CO₂ are fossil-fuel burning and carbon releases associated with land-use changes while the dominant sinks may involve a combination of oceanic uptake and uptake by terrestrial ecosystems^{1,2}. With one exception, these sources and sinks of CO₂ are stoichiometrically linked to sources and sinks of O₂. The exception is the uptake of CO₂ by the oceans which essentially proceeds through the reaction of dissolved CO₂ with carbonate ions and therefore does not involve O₂ (ref. 3). Thus, in principle, the difference between the decrease in atmospheric O₂, corrected for reaction stoichiometry, and the increase in atmospheric CO₂ can determine the uptake of CO₂ by physical dissolution in the oceans.

The acid-base chemistry of CO₂ also leads to qualitative differences in the fluxes of O₂ and CO₂ across the air-sea interface that are driven by marine biological processes. Marine photosynthesis and respiration produce comparable changes (in mol m⁻³) in dissolved O₂ and total dissolved inorganic carbon in sea water. Changes in dissolved CO₂ gas are ~15 times smaller, however, as determined by the chemical equilibrium between dissolved CO₂ gas and the other (dominant) forms of dissolved inorganic carbon, especially bicarbonate and carbonate ions^{3,24}. Because air-sea gas fluxes tend to be driven in proportion to the saturation anomalies in dissolved gas concentrations⁴, biologically induced air-sea fluxes of CO₂ are ~15 times smaller than fluxes of O₂ (ref. 5). Other important qualitative differences in the behaviour of CO₂ and O₂ arise because the equilibration timescale for dissolved CO₂ is ~15 times longer (~1 year) than that of dissolved O₂ (a few weeks)^{3,5}.

Observed variations in atmospheric oxygen

Previous attempts to measure changes in the percentage of

atmospheric O₂ in background air were thwarted by the high O₂ content of air^{6,7}. Here we report measurements using a new interferometric technique^{8,9} which overcomes that limitation.

We detect oxygen changes through changes in the O₂/N₂ ratio of air samples. Because atmospheric N₂ is believed to be much less variable than O₂, changes in the O₂/N₂ ratio should primarily reflect changes in O₂ (ref. 8). Changes in O₂/N₂ are reported as relative deviations from a reference

$$\delta(\text{O}_2/\text{N}_2) = \left(\frac{(\text{O}_2/\text{N}_2)_{\text{sample}}}{(\text{O}_2/\text{N}_2)_{\text{ref}}} - 1 \right) \quad (1)$$

We propose the convention of multiplying $\delta(\text{O}_2/\text{N}_2)$ by 10⁶ and expressing the result in units of 'per meg'. In these units 1/0.2095 = 4.8 per meg is equivalent to 1 p.p.m. by volume (p.p.m.v) because O₂ comprises 20.95% of air by volume⁷.

Our reference for the O₂/N₂ determinations is based on the average composition of the air derived from a suite of six high-pressure cylinders. The O₂/N₂ ratios of flask samples are determined first relative to a working gas. We compare the working gas frequently with a pair of secondary reference gases which, in turn, are compared at roughly six-month intervals with the six primary reference gases. These primary reference gases have shown detectable changes relative to each other but no systematic trends over a period of four years.

External precision, based on analysis of replicate flasks, is roughly ±4 per meg. Additional uncertainty must be allowed, however, for day-to-day variations in O₂/N₂ ratio of the air delivered from the reference gas cylinders. Allowing for this variability, we estimate the overall precision of the average concentration of the replicate flasks to be ±6 per meg relative to the long-term mean composition of the air delivered from the primary reference cylinders.

Results from samples collected at three sea-level sites, Alert, Ellesmere Island (82.5° N, 62.3° W), La Jolla, California (32.9° N, 117.3° W) and Cape Grim, Tasmania (40.7° S, 114.7° E) are shown in Fig. 1. Significant seasonal variations in $\delta(\text{O}_2/\text{N}_2)$ occur at all three sites. An interannual decrease in O₂/N₂ is clearly evident in the La Jolla data. As discussed below, the decrease rate is consistent with the global rate of O₂ consumption from the burning of fossil fuels. Oxygen depletion from burning fossil fuels is not a serious environmental concern because the changes are minuscule compared with the total O₂ abundance in air¹⁰.

Interpretation of seasonal variations

At the two Northern Hemisphere sites, the seasonal variations in O₂/N₂ are strongly anticorrelated with variations in CO₂. The amplitude of the O₂/N₂ cycle is larger than would be expected from the CO₂ variations. The CO₂ cycle is known to

be caused primarily by seasonal uptake and release of CO_2 by land plants and soils¹¹. Taking an average $\text{O}_2:\text{CO}_2$ exchange ratio for the formation and destruction of terrestrial organic matter of $-1.05:1$ (ref. 8), we estimate that $\delta(\text{O}_2/\text{N}_2)$ should increase by 5 per meg for each 1 p.p.m. decrease in CO_2 over the seasonal cycle. Instead, the observed ratio at the Northern Hemisphere sites is closer to 10 per meg per p.p.m. At Cape Grim in the Southern Hemisphere, the O_2/N_2 cycle is even more remarkable with a ~ 96 per meg seasonal variation in $\delta(\text{O}_2/\text{N}_2)$ occurring in the presence of a very small (~ 1 p.p.m.) variation in CO_2 .

In addition to exchanges with terrestrial ecosystems, there are two processes that can contribute significantly to seasonal variations in O_2/N_2 . First, the amount of O_2 that is dissolved in the upper ocean varies seasonally as the water warms and cools. Because O_2 in upper ocean equilibrates rapidly with the atmosphere, we can estimate the air-sea flux of O_2 associated with seasonal temperature changes from the relation⁵

$$F_{\text{O}_2} = -\frac{dC_{\text{eq}}}{dT} \frac{\dot{Q}}{C_P} \quad (2)$$

where F_{O_2} is the net O_2 flux (positive into the air) in $\text{mol m}^{-2} \text{s}^{-1}$,

dC_{eq}/dT is the temperature derivative of the O_2 solubility ($\text{mol m}^{-3} \text{K}^{-1}$), $\dot{Q}$ is the rate of change of heat storage in the upper ocean ($\text{J m}^{-2} \text{s}^{-1}$) and C_P is the heat capacity of sea water ($\text{J m}^{-3} \text{K}^{-1}$). The effect of these fluxes on the atmospheric O_2/N_2 ratio is reduced by $\sim 44\%$ due to a thermally driven flux of N_2 (ref. 5) which we can estimate by applying equation (2) to N_2 .

Second, there is net production of O_2 in the euphotic zone (roughly the upper 100 m of the ocean), where, on average, gross photosynthesis exceeds gross respiration, and there is a net consumption of O_2 below the euphotic zone by respiration. Both the net O_2 production rate in the euphotic zone and the rate of vertical mixing between the euphotic zone and deeper waters undergo large seasonal variations. At middle and high latitudes the highest production rates tend to occur in the spring and summer when the waters above 100 m are well stratified^{24,25}. A considerable fraction of the O_2 produced at this time escapes into the atmosphere. Comparable amounts of O_2 are removed from the atmosphere in the autumn and winter when stratification breaks down and oxygen-depleted waters from below 100 m depth are brought back into contact with the surface. The air-sea O_2 fluxes associated with this seasonal cycle are linked to the rate at which organic material is produced and exported from the euphotic zone^{5,12}.

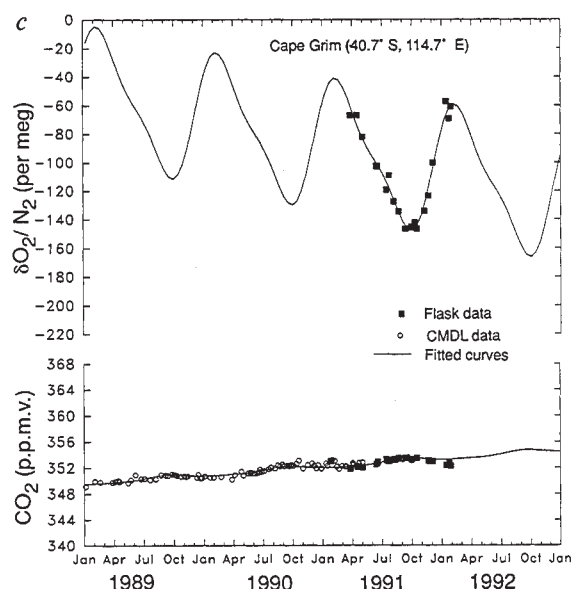
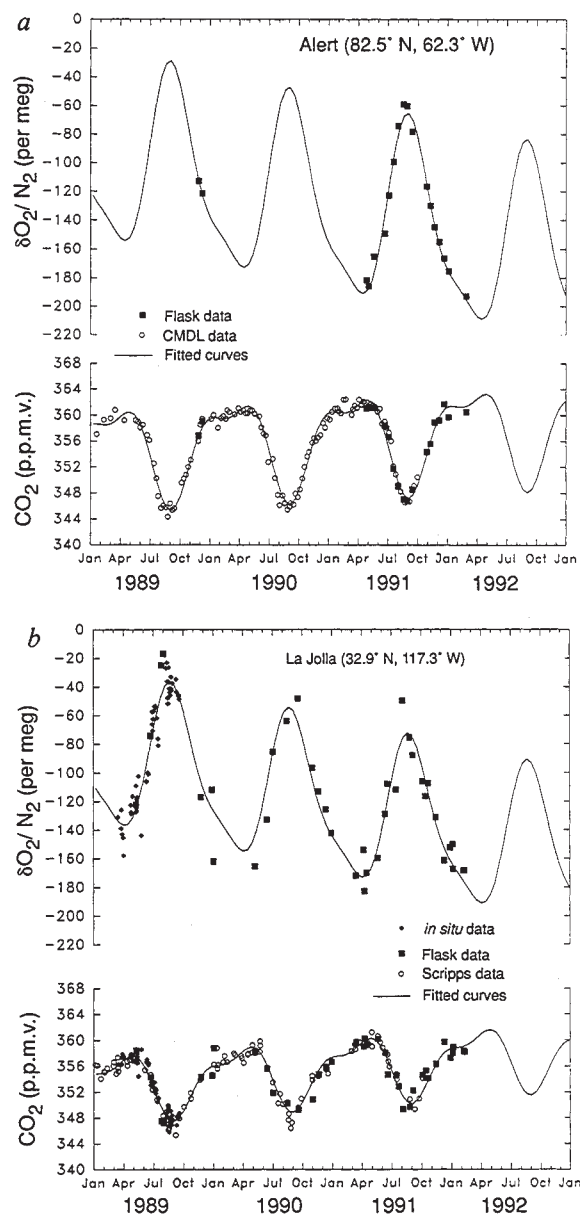


FIG. 1 Measurements of $\delta(\text{O}_2/\text{N}_2)$ and CO_2 mole fraction at (a) Alert, (b) La Jolla and (c) Cape Grim. The axes are scaled (5 per meg $\equiv$ 1 p.p.m.) so that changes in $\delta(\text{O}_2/\text{N}_2)$ and CO_2 are directly comparable on a mol O_2 to mol CO_2 basis. ■, Averages of flasks taken on a given day; ◆, daily averages for background conditions using the continuous *in situ* measurement method⁹. ○, Supplementary CO_2 data from programmes at the NOAA Climate Monitoring and Diagnostics Laboratory and the Scripps Institution of Oceanography. The curves are least-squares fits to the function $X = a + bt + c \sin(2\pi t) + d \cos(2\pi t) + e \sin(4\pi t) + f \cos(4\pi t)$, where t is in years. For the Cape Grim and Alert O_2/N_2 data, the trend coefficient b is forced to have the same value as for the La Jolla data; all other curves are six-parameter fits. Samples are collected in 5-litre glass flasks equipped with a pair of glass-bore stopcocks sealed with Viton O-rings. Flasks are generally collected in triplicate. Sampling times are chosen so that the air is at or near background composition as indicated by other tracers (such as CO_2 or radon). The flasks are exposed at the sampling sites by flushing them with cryogenically ($< -50^\circ \text{C}$) dried air at ambient pressure. Air samples are further dried with two cold traps ($< -80^\circ \text{C}$) before analysis. Samples are analysed for CO_2 concentration using an infrared analyser (Siemens Ultramatt 3) which is calibrated against CO_2 standards prepared at the Scripps Institution of Oceanography. The O_2/N_2 measurements are corrected for the small interfering effect of CO_2 on the refractivity ratio. Interferences of other trace gases may contribute to an offset of several per meg to the O_2/N_2 determinations, although the variability in this offset is believed to be smaller than the imprecision of the measurements.

Marine photosynthesis and respiration as well as changes in surface water temperature also produce variations in the CO_2 partial pressure of surface waters. The air-sea CO_2 exchange driven biologically is heavily suppressed, however, by the inorganic chemistry of CO_2 in sea water, as mentioned above. The thermal effects on p_{CO_2} are comparable in magnitude and generally opposite in sign to the biological effects, thus leading to even smaller seasonal air-sea CO_2 fluxes than would be expected from either marine biology processes or thermal processes alone¹³. In contrast, the thermal effects and biological effects on O_2 tend to reinforce each other as both lead to higher O_2 supersaturations in the warmer months¹².

In the Northern Hemisphere, we can use the seasonal variations in CO_2 , which largely reflect terrestrial exchange, to correct for the effects of terrestrial exchange on the O_2/N_2 variations, thus deriving the residual O_2/N_2 variations attributable to the oceanic exchanges alone (Fig. 2). In the Southern Hemisphere, the lack of a large seasonal variation in CO_2 indicates that the O_2/N_2 variations are almost entirely of marine origin.

The amplitude of the oceanic component of the O_2/N_2 variations within each hemisphere is closely related to the total seasonal net outgassing (SNO) of O_2 in that hemisphere. Here SNO is defined as the areally and temporally integrated flux of O_2 from the ocean into the atmosphere where the time integration extends over the seasonal period when the areally integrated flux is directed into the atmosphere. The close relationship between SNO and the seasonal O_2/N_2 variations follows because the timescale for seasonal variations ($t_{\text{season}} \approx 1 \text{ year}/2\pi = 66 \text{ days}$) is short relative to the timescale for atmospheric inter-hemispheric exchange ($t_{\text{exch}} \approx 1.1 \text{ years}^2$) but longer than the timescale for tropospheric mixing within each hemisphere ($t_{\text{hem}} \approx 30 \text{ to } 60 \text{ days}$). Here we construct preliminary estimates

of SNO for the Northern and Southern Hemispheres (Table 1) by assuming that the seasonal amplitudes in O_2/N_2 at the surface are equivalent to what would be achieved if the surface O_2 fluxes were homogeneously diluted into an air mass equal to 50% of the air mass in each hemisphere. This level of atmospheric mixing is consistent with mixing rates for atmospheric CO_2 in the Northern Hemisphere based on recent two- and three-dimensional transport models calibrated with other tracers^{11,14}.

The estimate of SNO derived in this way includes both biological and thermal components. The thermal component, however, includes only the excess O_2 flux not compensated for by the thermal N_2 flux because only the uncompensated O_2 flux can change the observed O_2/N_2 ratio. The uncompensated thermal component can be estimated independently by applying equation (2) to O_2 and N_2 , and thus the biological component can be derived by subtraction. The results (Table 1) indicate that uncompensated thermal degassing accounts for only ~15% of the oceanic component of the O_2/N_2 variations; the remaining 85% is evidently of marine biological origin.

What do these oxygen fluxes tell us about the rates of biological productivity in the ocean? To answer this question it is useful to introduce the concept of net euphotic zone production (NEZP), defined as the net community production (gross production of autotrophs less the gross respiration of autotrophs and heterotrophs) integrated over the depth of the euphotic zone. NEZP may be defined with respect to either organic carbon or O_2 . If we accept that net community production of carbon or O_2 requires inputs of 'new' nutrients, then NEZP may be roughly equated with the concept of 'new' production as given in ref. 15. We prefer to express our results in terms of NEZP because the concept of 'new' production of organic matter or O_2 becomes ambiguous whenever organic matter is formed with varying C:N or O_2 :N ratios, or whenever significant nitrification occurs in the euphotic zone.

Areally and annually integrated NEZP of O_2 can be related

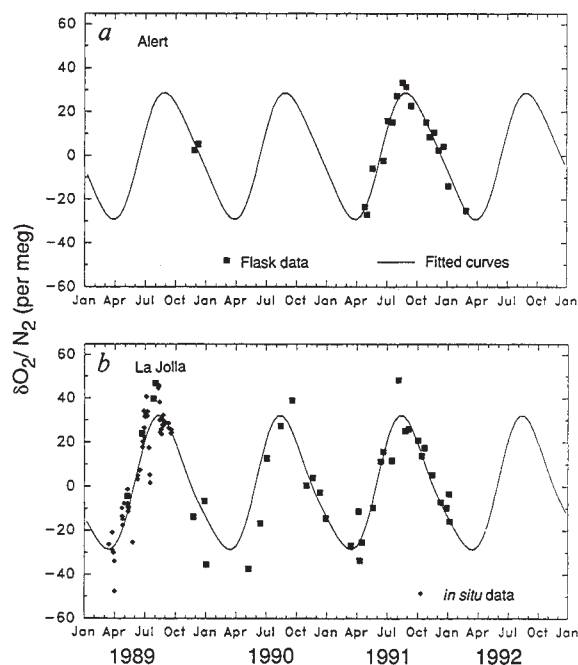


FIG. 2 Detrended oceanic components of the O_2/N_2 variations at Alert (a) and La Jolla (b). The oceanic component was derived according to $\delta(\text{O}_2/\text{N}_2)_{\text{oceanic}} = \delta(\text{O}_2/\text{N}_2) + (1.05/0.2095)X_{\text{CO}_2}$ where 1.05 is the adopted $-\text{O}_2:\text{CO}_2$ ratio for exchanges with terrestrial ecosystems, 0.2095 is the O_2 mole fraction of dry air (1/0.2095 is the appropriate conversion factor between per meg units and mole fraction units) and X_{CO_2} is the CO_2 mole fraction in p.p.m. The curve is a least-squares fit of the $\delta(\text{O}_2/\text{N}_2)_{\text{oceanic}}$ data to the same functional form as was used in Fig. 1. The $\delta(\text{O}_2/\text{N}_2)_{\text{oceanic}}$ data and the fitted curve were detrended by subtracting the function $a + bt$, where a and b are the coefficients from the six-parameter fit.

TABLE 1 Components of seasonal variations in oxygen

	Northern Hemisphere	Southern Hemisphere
Total seasonal amplitude	120 per meg*	95 per meg†
Oceanic component of seasonal amplitude	59 per meg‡	85 per meg§
Total SNO	$5.5 \times 10^{14} \text{ mol O}_2$	$7.8 \times 10^{14} \text{ mol O}_2$
Excess thermal component of SNO¶	$0.84 \times 10^{14} \text{ mol O}_2$	$0.97 \times 10^{14} \text{ mol O}_2$
Biological component of SNO#	$4.6 \times 10^{14} \text{ mol O}_2$	$6.8 \times 10^{14} \text{ mol O}_2$

* Estimated from the average of the peak-to-peak amplitudes for the fitted curves for the La Jolla and Alert data shown in Fig. 1.

† Estimated from the peak-to-peak amplitude of the fitted curve for the Cape Grim data in Fig. 1.

‡ Estimated from the average of the peak-to-peak amplitudes for the fitted curves for the La Jolla and Alert data shown in Fig. 2.

§ Assumes that the terrestrial exchanges can account for 10 per meg of the seasonal variations at Cape Grim, on the basis of a modelling study for CO_2 at 41°S , 175°E by Heimann *et al.*¹¹.

|| Calculated assuming that the surface fluxes are effectively diluted into a volume containing $4.4 \times 10^{19} \text{ mol}$ of dry air (50% of the dry air mass in the hemisphere).

¶ Estimated from equation (2) using ocean heat-storage data of Samuels and Cox²¹ and solubility data of Weiss²² (see ref. 5). The reported O_2 flux has been reduced by 46% to account for the compensating flux of N_2 . In other words, the reported flux accounts only for the excess O_2 flux not compensated by N_2 .

Calculated by subtracting the uncompensated thermal component of SNO from total SNO. The biological and thermal components of the SNO should be largely in phase with each other⁵.

to biological SNO as follows

$$\text{SNO} = g \text{ NEZP} \quad (3)$$

where g is a dimensionless number between 0 and 1.0. g is less than unity because only a fraction of the net O_2 produced throughout the year is reflected in the spring-summer pulse of O_2 injected into the atmosphere: some O_2 is produced in autumn and winter when the net flux is directed into the ocean, and some of the summertime O_2 production is exported into the oxygen-depleted water below the euphotic zone rather than into the atmosphere. In general, we expect middle and high latitudes to dominate the hemispheric SNO, whereas low, middle and high latitudes should all contribute significantly to hemispheric NEZP. We expect that upwelling regions, where O_2 fluxes are generally into the ocean throughout the year but where productivity is high, will reduce hemispheric SNO but make positive contributions to hemispheric NEZP. At present, the uncertainty in g makes it difficult to relate our oxygen signals directly to NEZP. For preliminary estimates, we will adopt $g = 0.5$ for the hemispherically averaged fluxes. This value seems plausible considering that the hemispheric average must include low-latitude waters, where seasonal outgassing is minimal, as well as middle- and high-latitude waters where a significant fraction, probably greater than 50%, of the annual net O_2 production is reflected in the seasonal outgassing^{12,16}.

Taking $g = 0.5$, the above estimates for SNO lead to estimates of 9.2×10^{14} and 13.6×10^{14} mol O_2 yr⁻¹ for NEZP for the Northern and Southern Hemispheres, respectively, or 22.8×10^{14} mol O_2 for the world as a whole. Taking an O_2 :C ratio of -1.4:1 (ref. 17) for marine organic matter, we obtain an estimate of 16×10^{14} mol C yr⁻¹ (19 gigatonne C yr⁻¹) for global NEZP of organic carbon, or 4 mol C m⁻² yr⁻¹ for the global average. This result is higher than recent estimates of global new production (4–17 Gt C yr⁻¹)^{18–20} but generally agrees with local estimates derived from observations of O_2 in the upper ocean¹².

The preliminary nature of this analysis must be emphasized: uncertainties in estimated productivities are probably a factor of two or more. Measurements from additional stations are needed to characterize more accurately the large-scale variability in O_2/N_2 . More accurate account of atmospheric transport and of the relation between oxygen fluxes and marine productivity is also needed (see also ref. 5). Nevertheless, these results illustrate the contribution that atmospheric O_2 measurements can make to studies of biological activity in the ocean. Measurements of the seasonal variations in O_2/N_2 extending over many years could provide an index of interannual variations in marine productivity on large spatial scales. Such data would be valuable for monitoring the sensitivity of marine productivity to climate change, decreasing stratospheric ozone, eutrophication or other perturbations.

Interpretation of interannual decrease

Schematically, we can represent the global budget for atmospheric CO_2 according to

$$\Delta\text{CO}_2 = F + C - O + B \quad (4)$$

where ΔCO_2 is the annual averaged change in atmospheric CO_2 , F is the source of CO_2 from burning fossil fuels, C is the CO_2 source from cement manufacturing, O is the oceanic CO_2 sink, and B is the net source of CO_2 from terrestrial ecosystems (B can be positive or negative), all in units of mol yr⁻¹. Likewise, we can represent the budget for atmospheric oxygen according to

$$\Delta\text{O}_2 = -F - H - \alpha_B B \quad (5)$$

where ΔO_2 is the change in atmospheric oxygen, H is the O_2 sink owing to the oxidation of elements other than carbon (predominantly hydrogen) in fossil fuels and α_B represents the O_2 :C exchange ratio for terrestrial biomass. The oceans do not significantly buffer the decrease in atmospheric O_2 because of the low solubility of O_2 in sea water⁸.

TABLE 2 Terms in global O_2 and CO_2 budgets

Symbol	Description	Value around 1990 (10^{14} mol yr ⁻¹)
Observed quantities		
$-\Delta\text{O}_2$	Measured O_2 depletion rate*	6.7 ± 1.7
ΔCO_2	Measured CO_2 increase†	2.3 ± 0.3
F	Fossil-fuel CO_2 source‡	4.85 ± 0.32
H	Fossil-fuel O_2 sink from elements other than carbon§	2.01 ± 0.17
$F + H$	Total fossil-fuel O_2 sink§	6.86 ± 0.41
C	CO_2 source from cement manufacturing¶	0.12 ± 0.01
Derived quantities		
B	Net terrestrial carbon source¶¶	-0.2 ± 1.7
O	Net oceanic CO_2 sink¶¶	2.5 ± 1.7

* Derived by multiplying the total number of moles of O_2 in the atmosphere (3.705×10^{19} (ref. 8)) by the coefficient ($b = -18.2$ per meg yr⁻¹) of the fitted curve to the La Jolla data in Fig. 1. The standard error in b estimated from the formal covariance matrix of the fit is ± 1.8 per meg yr⁻¹, but we adopt an uncertainty of ± 5 per meg yr⁻¹ based on a subjective assessment of possible drift in the reference gases.

† Derived from the coefficient ($b = 1.25 \pm 0.14$ p.p.m. yr⁻¹) of the curve fitted to the La Jolla CO_2 data in Fig. 1. The uncertainty in b is estimated from the formal covariance matrix of the fit assuming that the standard error in individual data points is equal to the mean square of the residuals to the fit.

‡ Using CO_2 production figures for 1989 from ref. 23.

§ From ref. 8 using fossil-fuel production figures for 1989 from ref. 23.

¶ Derived using $B = -(1/\alpha_B)(\Delta\text{O}_2 + F + H)$ from equation (5) using $\alpha_B = 1.05$ (ref. 8).

¶¶ Derived using equation (5).

Adding equations (4) and (5) and solving for O yields

$$O = -(\Delta\text{O}_2 + H) - (\Delta\text{CO}_2 - C) - (\alpha_B - 1)B \quad (6)$$

The last term on the right-hand side of equation (6) can be evaluated by solving equation (5) for B , although this term is virtually negligible as $\alpha_B \approx 1$.

Once sufficiently long time series for oxygen are available, it will be possible to estimate oceanic uptake of CO_2 using equation (6). Although the existing records are too short to yield definitive results, we can nevertheless use these data to illustrate the method. We have estimated (Table 2) the terms in equations (4) to (6) using the data from La Jolla to estimate global trends in O_2 and CO_2 . The global O_2 depletion rate inferred from the La Jolla data agrees well with the trend expected from the burning of fossil fuels, implying that the net terrestrial carbon source is zero within the uncertainties. The data suggest that the oceans were a sink of $2.5 \pm 1.7 \times 10^{14}$ mol C yr⁻¹ (3.0 ± 2.0 Gt C yr⁻¹) over the period from 1989 to 1991. This estimate of the oceanic sink agrees, within rather wide error margins, with estimates derived from ocean models calibrated with bomb ^{14}C data¹. The largest uncertainties in the estimated oceanic uptake are contributed by the trend in O_2/N_2 , and this uncertainty will decrease as longer time series are obtained.

Accurate determination of the O_2 depletion rate from measurement time series will require very stable standards. In this respect the present procedure, which relies on the stability of O_2/N_2 in compressed air cylinders, may not be adequate. To insure against the possibility of parallel drift in O_2/N_2 in reference cylinders, new methods of absolute standardization of O_2/N_2 mixtures are needed. □

Received 13 May; accepted 23 July 1992.

1. Keeling, C. D. et al. in *Aspects of Climate Variability in the Pacific and the Western Americas*, *Geophys. Monogr.* 55 (ed. Peterson, D. H.) 165–236 (American Geophysical Union, Washington DC, 1989).

2. Tans, P. P., Fung, I. Y. & Takahashi, T. *Science* **247**, 1431–1438 (1990).

3. Broecker, W. S. & Peng, T.-H. *Tracers in the Sea* (Lamont-Doherty Geological Observatory, New York, 1982).

4. Liss, P. S. & Merliat, L. in *The Role of Air-Sea Exchange in Geochemical Cycling* (ed. P. Buat-Menard) 113-127 (Reidel, Dordrecht, 1986).
5. Keeling, R. F., Bender, M. M., Najjar, R. G. & Tans, P. P. *Glob. biogeochem. Cycles* (submitted).
6. Benedict, F. G. *The Composition of the Atmosphere with Special Reference to its Oxygen Content* (Carnegie Institution of Washington, Washington DC, 1912).
7. Machta, L. & Hughes, E. *Science* **168**, 1582-1584 (1970).
8. Keeling, R. F. thesis, Harvard Univ. (1988).
9. Keeling, R. F. *J. atmos. Chem.* **7**, 153-176 (1988).
10. Broecker, W. S. *Science* **168**, 1537-1538 (1970).
11. Heimann, M., Keeling, C. D. & Tucker, C. J. in *Aspects of Climate Variability in the Pacific and the Western Americas*, *Geophys. Monogr.* **55** (ed. Peterson, D. H.) 277-303 (American Geophysical Union, Washington DC, 1989).
12. Jenkins, W. J. & Goldman, J. C. *J. mar. Res.* **43**, 465-491 (1985).
13. Takahashi, T., Goddard, J., Chipman, D. W., Sutherland, S. C. & Mathieu, G. *Final Tech. Rep. to DOE under contract 19X-SC428C*, 1-111 (Lamont Doherty Geological Observatory, New York, 1991).
14. Tans, P. P., Conway, T. J. & Nakazawa, T. *J. geophys. Res.* **94**, 5151-5172 (1989).
15. Dugdale, R. C. & Goering, J. J. *Limnol. Oceanogr.* **23**, 196-206 (1967).
16. Emerson, S. *J. geophys. Res.* **92**, 6535-6544 (1987).
17. Laws, E. A. *Deep-Sea Res.* **38**, 143-167 (1991).
18. Eppley, R. & Peterson, B. J. *Nature* **282**, 677-680 (1979).
19. Martin, J. M., Knauer, G. A., Karl, D. M. & Broenkow, W. W. *Deep-Sea Res.* **34**, 267-285 (1987).
20. Najjar, R. G., Sarmiento, J. L. & Toggweiler, J. R. *Glob. biogeochem. Cycles* **6**, 45-76 (1992).
21. Samuels, B. L. & Cox, M. *Ocean Model.* **75**, 1-3 (1987).
22. Weiss, R. F. *Deep-Sea Res.* **17**, 721-735 (1970).
23. Marland, G. & Boden, T. in *Trends 91: A Compendium of Data on Global Change*, *ORNL/CDIAC-46* (eds Boden, T. A., Sepanski, R. J. & Stoss, F. W.) **389** (CDIAC, Oak Ridge Tennessee, 1991).
24. Peng, T. H., Takahashi, T., Broecker, W. S. & Olafsson, J. *Tellus* **B39**, 439-458 (1987).
25. Asper, V. L., Deuser, W. G., Knauer, G. A. & Lohrenz, S. E. *Nature* **357**, 670-672 (1992).

ACKNOWLEDGEMENTS. We thank N. Trivet and V. Chorney of the Canadian Baseline Program, R. Francey of CSIRO, S. Wilson and L. Porter of the Cape Grim Baseline Air Pollution Station, and C. D. Keeling, A. Bollenbacher and D. Moss of the Scripps Institution of Oceanography for help with sampling; J. Anderson, P. McKinney and J. Firor for support; E. Atlas, D. Erickson, R. Najjar, M. Bender and U. Siegenthaler for helpful comments; and A. Manning for help with the figures. This work was supported by the US NSF; additional support for R.K. was provided by a postdoctoral fellowship in the Advanced Study Program at NCAR.

Crystal structures explain functional properties of two *E. coli* porins

S. W. Cowan, T. Schirmer, G. Rummel*, M. Steiert*, R. Ghosh*, R. A. Pauptit†, J. N. Jansonius‡ & J. P. Rosenbusch*

Department of Structural Biology, * Department of Microbiology, Biocentre, University of Basel, CH-4056 Basel, Switzerland

Porins form aqueous channels that aid the diffusion of small hydrophilic molecules across the outer membrane of Gram-negative bacteria. The crystal structures of matrix porin and phosphoporin both reveal trimers of identical subunits, each subunit consisting of a 16-stranded anti-parallel β -barrel containing a pore. A long loop inside the barrel contributes to a constriction of the channel where the charge distribution affects ion selectivity. The structures explain at the molecular level functional characteristics and their alterations by known mutations.

THE outer membrane of Gram-negative bacteria protects the cell from harmful agents such as proteases, bile salts, antibiotics, toxins and phages, and against drastic changes in osmotic pressure. Such a barrier must contain proteins that help the uptake and disposal of small hydrophilic compounds such as nutrients and waste products. In *Escherichia coli* these proteins include specific translocators and general diffusion pores whose expression depends on growth conditions. For example, the general diffusion pores matrix porin (OmpF) and osmoporin (OmpC), coded by the genes *ompF* and *ompC*, are regulated by osmotic pressure and temperature. Porins, which exhibit some specificity in addition to their general diffusion properties, include phosphoporin (PhoE), which is derepressed under phosphate starvation, and maltoporin (LamB), which is induced by maltose. Porins are extremely resistant to detergents and proteases, vital properties if cells are to survive in the intestinal tract. Several recent reviews cover porin structure and function¹⁻³. The OmpF, OmpC and PhoE porins are highly homologous (OmpF and PhoE share 63% sequence identity⁴). Although they have similar solute exclusion sizes ($M_r \sim 600$), other aspects of their channel characteristics differ. The OmpF and OmpC porins are weakly cation selective whereas PhoE is weakly anion selective⁵. When reconstituted into lipid bilayers, they show distinct channel conductances (0.8, 0.5 and 0.6 nS in 1M salt^{6,7}, respectively), and voltage gating behaviour⁸ (with respective threshold potentials of 90, 160 and 100 mV, above which the channels close).

The porins seem unusual for membrane proteins because their sequences are predominantly polar and contain no long hydrophobic segments like those found in membrane-spanning α -

helices of bacteriorhodopsin⁹ and the bacterial photosynthetic reaction centre¹⁰. In fact, spectroscopy revealed that the secondary structure of OmpF porin consists almost entirely of β -sheet¹¹. This porin was the first membrane protein amenable to analysis by X-ray crystallography^{12,13}. The recent structure determination to high resolution of a porin from *Rhodobacter capsulatus*¹⁴ revealed a 16-stranded anti-parallel β -barrel enclosing an aqueous pore. Despite the lack of sequence homology between this porin and the *E. coli* porins, it was possible to confirm the barrel motif for three *E. coli* porins¹⁵ using molecular replacement methods¹⁶ at 6.0 Å resolution. But the structures were too different to be determined by this method. Here we present the structure of OmpF porin solved by single isomorphous replacement at 2.4 Å resolution (Table 1), and that of PhoE porin solved by molecular replacement at 3.0 Å resolution using the OmpF model. Most functional and topological investigations have been done with *E. coli* porins because of their intensively studied genetics. Their unusual stability has been critical in obtaining a host of biochemical data relating to their properties^{1,2}. These data can now be readily correlated with the structures presented in this paper. A comparison of the OmpF and PhoE porins reveals the structural basis for their distinct pore characteristics in atomic detail.

Description of the structures

The channel-forming motif of both porins is a 16-stranded anti-parallel β -barrel with (+1)₁₅ topology¹⁷ (Fig. 1a, b) and a shear number¹⁸ of 20. The tilt of the strands relative to the barrel axis varies between 35° and 50°. The amino and carboxy termini are linked by a salt-bridge within the 16th β -strand, thus forming a pseudo-cyclic structure. Short β -hairpin turns define one end of the barrel (the smooth end) whereas long irregular loops are found at the other end of the barrel (the rough end). The loops are tightly packed and extend towards the barrel axis. The pore

† Present address: ICI Pharmaceuticals, Mereside, Alderley Park, Macclesfield, Cheshire SK10 4TG, UK.

‡ To whom correspondence should be addressed.

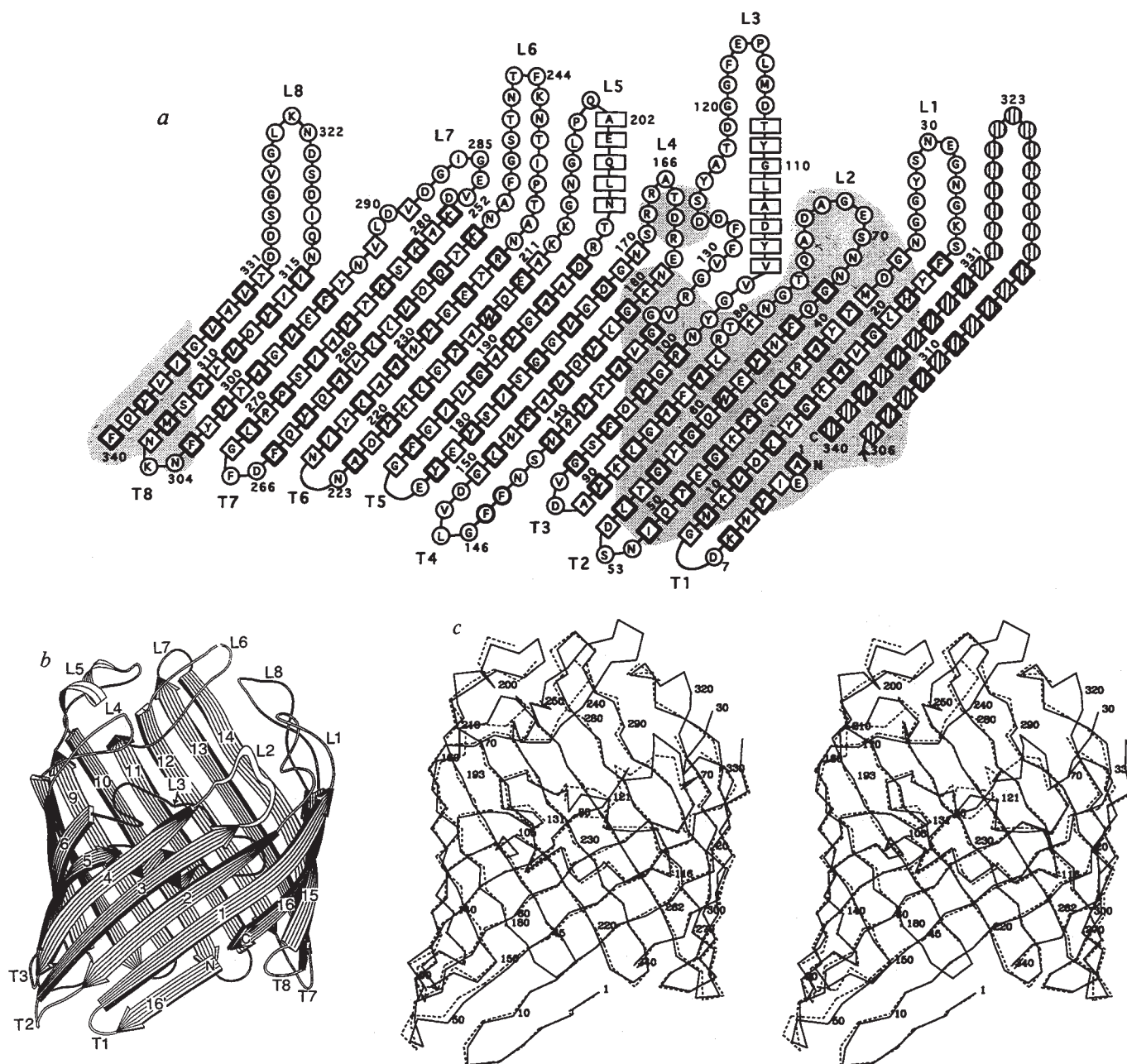


FIG. 1 Structures of the OmpF and PhoE porin monomers. *a*, Topology of OmpF, with amino-acid sequence in one-letter code. The view is from outside the 16-stranded antiparallel β -barrel, which is unrolled. The last two β -strands are repeated on the right-hand side (hatched) to emphasize the cyclic nature of the structure. Secondary structural elements are indicated (bold if their side chains are external) by diamonds for barrel β -strands, rectangles for α -helices and circles for turns and loops. The subunit contact regions are shaded. *b*, RIBBON diagram⁵⁸ of OmpF. Arrows represent β -strands and are labelled 1 to 16 starting from the strand after the first short turn (the prefix β is used in the text). The short β -strand at the N terminus continues the C-terminal strand β 16 and is therefore called β 16'. The long loops are denoted L1 to L8, the short turns at the other end T1 to T8. Loop L2 protrudes towards the viewer. Loop L3 (black) folds inside

lies off-centre and at an angle of 16° to the barrel axis. One of the long loops (L3) folds into the barrel leaving a gap in the wall between strands β 4 and β 7. This loop, which contains one-and-a-half turns of α -helix, constricts the size of the pore at about half the height of the barrel. The differences in backbone folding between OmpF and PhoE (Fig. 1c), consequences of evolutionary insertions and deletions between the two proteins, are confined to the loop regions and a single short turn. A local

change in the conformation of the backbone caused by the insertion of Ser 121A in L3 allows room for the side chain of Lys 131 of PhoE (glycine in OmpF). The barrel structures of the two proteins are very similar. The 190 barrel $C\alpha$ atoms can be superimposed with a root mean square (r.m.s.) deviation of 0.54 Å. The major contribution to trimer formation is hydrophobic (the regions involved in subunit interactions are shaded in

the barrel. *c*, Stereo representation of superimposed $C\alpha$ -tracings of OmpF (solid line) and PhoE (dashed line); the r.m.s. deviations for 279 $C\alpha$ positions is 0.74 Å. The barrel structure and most of the internal loop L3 are very similar. Major differences exist in the external loops at the rough end of the barrel (top) where most insertions and deletions between the two porins occur. Loop L2, facing the viewer, latches onto an adjacent monomer in the trimer. The OmpF residue numbering scheme has been adopted for PhoE causing gaps at residues 27–31, 75–76, and 242–247, and additions at residues 121A (after 121), 183A and 183B. The following residues could not be modelled because of weak or nonexistent density; OmpF: 26–29 (loop L1); PhoE: 25–33 (L1), 181–184 (turn T5), 198–199 (L5), 204 (L5), 249 (L6) and 319–321 (L8).

change in the conformation of the backbone caused by the insertion of Ser 121A in L3 allows room for the side chain of Lys 131 of PhoE (glycine in OmpF). The barrel structures of the two proteins are very similar. The 190 barrel $C\alpha$ atoms can be superimposed with a root mean square (r.m.s.) deviation of 0.54 Å.

The major contribution to trimer formation is hydrophobic (the regions involved in subunit interactions are shaded in

FIG. 2 The structure of the trimer. *a*, Stereodigram of the α -carbon trace of OmpF. The view is about 30° skew to the 3-fold axis which is shown in the centre of the trimer. The monomers are coloured differently to highlight the latching of loop L2 between them (adjacent to the 3-fold axis). The orthogonal packing of β -sheets is visible between the blue and the green subunits. *b*, Stereodigram of the OmpF trimer with atoms represented by spheres corresponding to their van der Waals radii. The view is perpendicular to the triad (vertical line) and the crevice between two monomers can be seen in the centre. The atoms are: carbon, yellow; nitrogen, blue; and oxygen, red. All carbon atoms in aromatic residues are coloured white to emphasize the boundaries of the hydrophobic band they delimit. The width of the hydrophobic band indicated on the scale on the right is measured between the hydroxyl groups of tyrosines in the aromatic rings. A patch of negative residues which may interact through divalent cations with lipid headgroups in the membrane can be seen just below the phenyl ring on top.

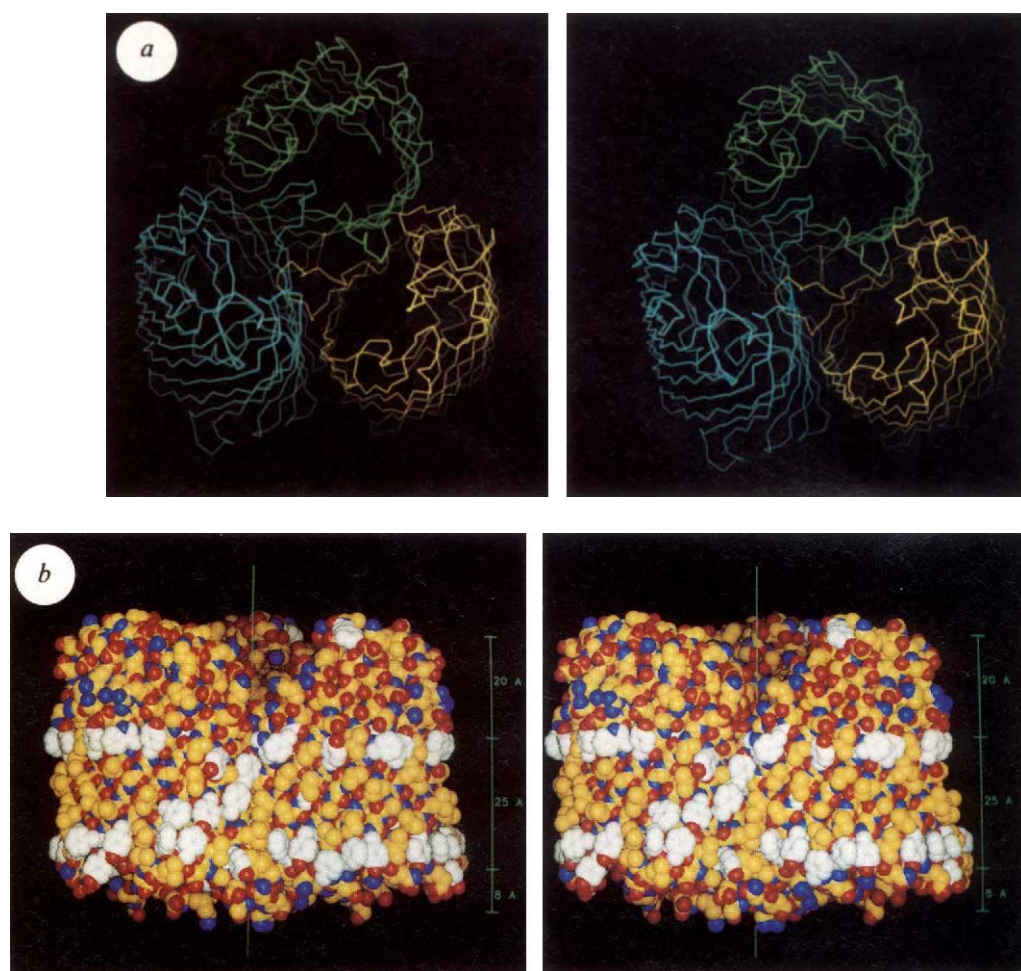
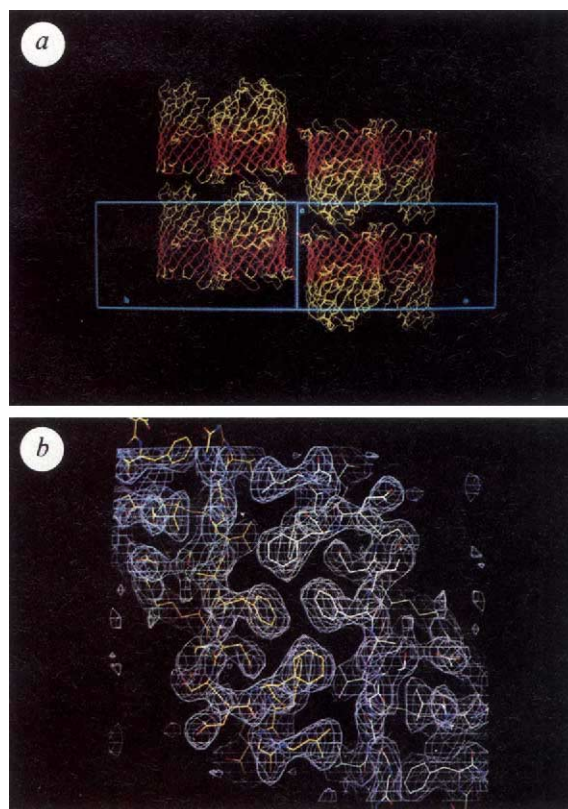


FIG. 3 The crystal packing of OmpF. *a*, The C α -tracings of four trimers (two unit cells) are coloured red where they represent part of the hydrophobic band which would be located in the core of the membrane. The unit cell is drawn in blue with the cell axes labelled. The view is perpendicular to *c* along a crystallographic 2-fold axis (a diagonal in the *a-b*-plane). Adjacent columns of trimers running in opposite directions stack with their hydrophobic bands in register forming two-dimensional layers in the *a-b*-plane. Few hydrophilic contacts exist between these layers. Owing to the opposite directions of adjacent trimers and the differing sizes of their two hydrophilic regions, the layers stack in an interlocking fashion in the third dimension. This is presumably why this crystal form is better ordered than a hexagonal crystal form where the trimers are all orientated in the same direction¹³. The solvent content of 58% is low for membrane protein crystals as it is within the range usually observed for soluble proteins. *b*, Atomic detail of the residues involved in hydrophobic contacts between trimers in the crystals of OmpF. This window is an enlargement of the area around the 2-fold axis halfway along *c* in Fig. 3*a*. In one molecule the bonds are drawn in yellow and in the other white. The Fourier map calculated with model phases and $2|F_o| - |F_c|$ coefficients is contoured at 1σ . The packing in the PhoE crystals is similar, but the side-chains involved in hydrophobic contacts are different.



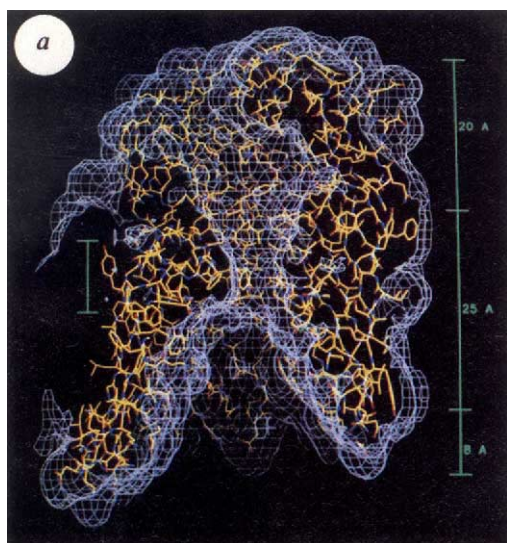
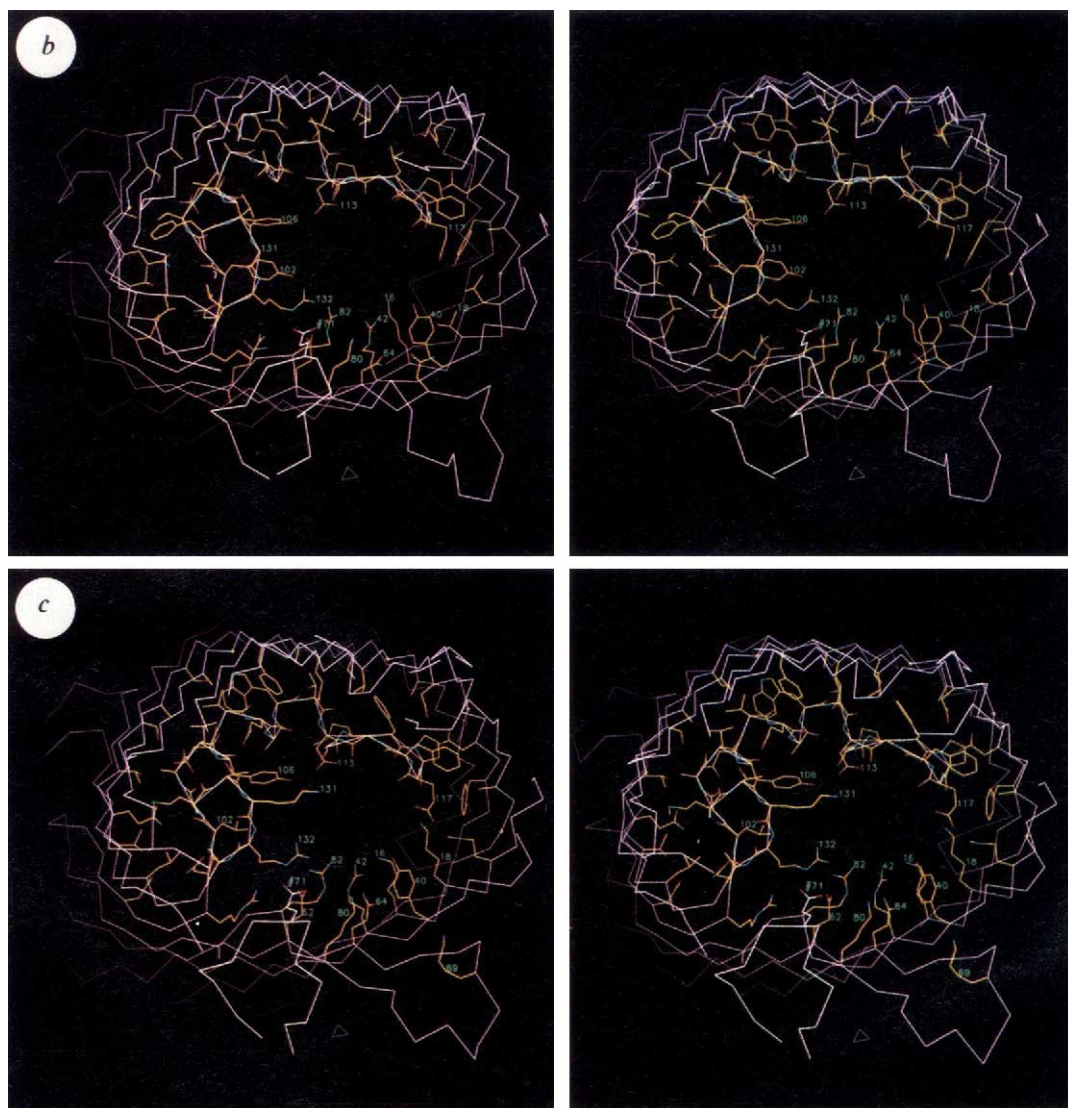


Fig. 1a). Owing to the tilting of the β -strands, the β -sheets at the subunit interface stack in an orthogonal manner (Fig. 2a). Aromatic residues from the barrel walls, which face glycine or alanine residues on the opposite walls, interdigitate. Owing to the tight packing of large hydrophobic residues along the symmetry axis of the trimer, there is no room for water in this interface, except near the surfaces, where several water molecules are seen. Loop L2 contributes to the integrity of the trimer by numerous hydrophilic interactions. This loop reaches between L2* and L4* of an adjacent monomer (the asterisk denotes structural elements from a neighbouring subunit), thus filling the gap left by L3*. In both porins, L2 forms hydrogen-bonds with residues from L2*, L3* and L4* and strong salt-bridges to the arginines in L3*. The salt-bridge connecting the N and C termini is located in the monomer-monomer interface and is not in contact with the lipid phase of the membrane. Deletion or mutation of the C-terminal residue, which is a highly conserved phenylalanine in porins, lowers trimer stability¹⁹, whereas deletion of the 16th β -strand abolishes trimer formation²⁰. The combination of hydrophobic and hydrophilic subunit interactions, which involve 35% of the surface of each isolated

FIG. 4 The pore structure. *a*, Solvent-accessible surface of the OmpF pore (probe radius used: 1.4 Å). The longitudinal section through the right-hand monomer in Fig. 2b shows residues from the internal loop L3 at the right and arginines from the inner barrel wall at the left. The axial length of the constriction zone (~9 Å) is indicated by the short vertical bar. The scale shows the relative height of the hydrophobic band as in Fig. 2b. The inner barrel wall is ~30 Å high, indicating that the channels do not merge on traversing the membrane. The barrel has an elliptical cross-section of 27 × 38 Å between main-chain atoms in the barrel walls. Pore dimensions: between 11 and 19 Å, at the mouth; 11 × 7 Å, in constriction zone; 22 × 15 Å, at smooth end. *b*, *c*, Stereodiagram of the constriction zones in the pores of OmpF and PhoE, respectively. The view is perpendicular to that in *a* along the trimer axis. The C α -tracing is represented by a mauve line. At the bottom the loop L2 protrudes from the barrel, near the triangle representing the 3-fold axis. L2* from a neighbouring monomer (white) protrudes into the pore. Residues are depicted in a slab about 10 Å thick. Residues involved in the packing of L3 against the barrel wall are at the top and left inside the barrel. Part of L3 is omitted for clarity. The tip of L3 consists of the conserved residues 116–120 (PEFGG). Arginine and lysine residues protrude from the barrel wall close to the 3-fold axis. Asp 113 and Glu 117 and numerous carbonyl oxygens from L3 contribute to the



constriction from the opposite side of the pore. The PhoE pore is more constricted than that of OmpF because of the conformation of Glu 117 and the substitution of Lys 131 for Gly. Other differences are Glu 64 (Asn in OmpF) and Lys 18 (Val in OmpF).

TABLE 1 Crystallographic data

(a) Data collection

Protein	Data set	Resolution (Å)	Number	Reflections (% of possible)	R_{merge} (%)	ΔF (%)
OmpF	Native	FAST Tris	3.0	8,653	98	7.0
	Native	FAST TEA	3.0	7,854	90	5.4
	Native	IP Tris	2.4	16,989	96	7.7
	Se-Met	FAST Tris	2.8	9,720	87	6.9
	PTEN	FAST TEA	3.0	7,800	90	5.2
PhoE	Native	FAST	3.0	8,685	95	9.0
						14.3

(b) OmpF phasing statistics

Heavy-atom parameters

Site	Relative occupancy	x	y	z	R_{centric}	40.0%
1	1.00	0.15	0.54	0.83	Mean figure of merit	
2	0.60	0.18	0.42	0.50	(15–3.0 Å)	0.38
3	0.45	0.07	0.54	0.79	F_H/E	1.65

(c) Refinement statistics

Protein	Completeness of model	Water molecules	R (%)	r.m.s. deviations
	Residues	Atoms	(for all data)	Bonds (Å)
OmpF	336/340	2,621/2,626	36	20.8 (8–2.4 Å)
PhoE	313/330	2,339/2,602	0	22.2 (8–3.0 Å)
				Angles (°)
				0.016
				0.021
				3.1
				3.9

Both proteins crystallize in space-group $P321$ with cell constants $a=b=118.5$, $c=52.7$ Å (OmpF) and $a=b=119.9$, $c=51.9$ Å (PhoE); solvent content ~58%. FAST, data collected on an Enraf-Nonius area detector and processed using MADNES⁴⁸; IP, image plate data collected and processed at EMBL, Hamburg; TEA, triethylamine; PTEN, platinum-ethylenediamine-dichloride. METHODS. OmpF and PhoE were overexpressed, purified and crystallized as described previously^{49,50}. Heavy-atom derivative searches were done with OmpF crystals in both the Tris-buffered mother liquor (pH 8.0) and a TEA buffer (pH 9.0). Many compounds bound, but caused nonisomorphism or yielded centrosymmetric heavy-atom structures. A useful derivative was prepared by soaking crystals in a solution of 2 mM PTEN for 8 days. The heavy-atom parameters were refined and phases calculated using programs REFIN and PHASE from the CCP4 package⁵¹. The 3.0 Å SIR phases were improved by cycles of solvent flattening⁵², initially with an envelope calculated using 4 Å multiple isomorphous replacement data (not shown) at the 35% solvent level and later with envelopes calculated using the improved SIR phases up to the 50% solvent level. A model containing 95% of the atoms (R factor 48.6%, 8–3 Å resolution) was built⁵³ into the resulting map. Refinement proceeded using X-PLOR⁵⁴ with simulated annealing, initially at 3.0 Å, then at 2.4 Å resolution with restrained isotropic B-factors, and model building using the program 'O'⁵⁵. The PhoE structure was solved using programs POLARRFN and TFSGEN⁵¹, with a model of OmpF which had external loop regions deleted and side chains pruned to common atoms. Rigid-body refinement of the solution reduced the R factor from 46.5% (10–5 Å) to 41.5% (10–3.2 Å) and was followed by model building and conventional positional refinement with an overall B factor. The correctness of the OmpF model was confirmed as follows. A difference Fourier map calculated with data from seleno-methionine substituted OmpF ($|F_{\text{Se-deriv}} - F_{\text{native}}|$ coefficients, model phases) has peaks $>14\sigma$ at the sulphur positions of the three methionines in the OmpF model. Further evidence was the solution of the PhoE structure with the OmpF model and the presence of unbiased density in loop regions of PhoE. Heavy-atom derivative difference Fourier maps for OmpF and PhoE show that all heavy atoms are bound to methionines or histidines.

$$R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum I$$

$$\Delta F = \sum |F_{\text{deriv}} - F_{\text{native}}| / \sum F_{\text{native}}$$

$$F_H/E = \text{'phasing power'} = (\text{r.m.s. calculated } F_H) / (\text{r.m.s. error})$$

$$R_{\text{centric}} = \sum ||F_{\text{deriv}} \pm F_{\text{native}}| - F_H(\text{calc})| / \sum |F_{\text{deriv}} - F_{\text{native}}|$$

$$R = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum F_{\text{obs}}$$

monomer, may well explain why these porins cannot be dissociated in high concentrations of chaotropic agents or organic solvents other than phenol²¹.

The structures presented here correlate very well with projection maps obtained with unstained samples by electron microscopy (EM)^{22,23}. Weak electron density observed on the trimer axis in the maps of PhoE and OmpF led to the suggestion of a binding pocket for one molecule of lipopolysaccharide (LPS)²⁴. Because the X-ray structures show there is no room for LPS in the interface and a conformational change seems unlikely, the low density in this region in the EM maps is presumably due to the weak scattering power of the large hydrophobic side chains in the trimer interface. On shadowing, freeze-dried, reconstituted porin membranes revealed a protrusion on the trimer axis which was interpreted to be LPS²⁵. This could be due to the seeding of ice by the water molecules in that location. The X-ray structures reveal that the channels are well separated over a distance of ~35 Å (Fig. 2a) as was seen in three-dimensional EM maps of PhoE^{24,26} and OmpF²⁷. The negatively stained OmpF maps showed a merging of the channels after 30 Å. As shown here the channels do not actually merge, but the rough end of the trimer has protrusions defined by loops L4–L8 between which stain would accumulate. In contrast, the *Rb. capsulatus* porin structure²⁸ has a wide and deep vestibule so that the channels run separately over a distance of only 20 Å.

External surface

The outer surface of porin shows a pronounced partitioning with respect to hydrophilicity. A hydrophobic band with a width of 25 Å encircles the trimer (Fig. 2b). This width is compatible with estimates of the thickness of the nonpolar outer membrane core based on the lengths of its fatty acyl chains and X-ray studies of model membranes²⁹. The boundaries of the band are largely composed of phenylalanine residues and tyrosine residues with their hydroxyl groups towards the hydrophilic regions on either side of the band. The remainder of the band is composed of small to medium size aliphatic residues. The strict segregation of aliphatic and aromatic residues breaks down in the region where the monomers meet (the crevice). A similar distribution was observed in the structure of *Rb. capsulatus* porin³⁰ and less conspicuously in the arrangement of mainly tryptophan residues at the boundary between hydrophilic and hydrophobic zones of the reaction centre³¹ and bacteriorhodopsin³². This suggests that polarizable aromatic rings are energetically favoured at an interface between media of vastly different dielectric constants. In porin these aromatic residues may also be involved in stacking with sugar residues in LPS, similar to interactions between periplasmic binding proteins and their ligands³³. The hydrophilic parts of the external surface are quite distinct in size (Fig. 2b). Residues from the β -hairpin turns at

TABLE 2 Effects of point mutations and deletions on porins

Residue(s)*		Location		Effect†		Impaired phage/antibody binding	Ref.
Number	Type	OmpF	OmpC	Increased pore size	Altered selectivity		
18	Val	<u>Lys</u>	Asp	Pore, below constriction zone L1	±	—	(47)
34	Gly	<u>Lys</u>	Val		±		(47)
42	<u>Arg</u>	<u>Arg</u>	<u>Arg</u>	Pore, constriction zone	+	—	(37, 45)
69	Asn	<u>Lys</u>	Ser	L2, mouth of pore	+		(47)
82	<u>Arg</u>	<u>Arg</u>	<u>Arg</u>	Pore, constriction zone	+	—	(37, 45)
108–133	Δ			L3	+	—	(37, 34)
111–118			Δ	L3	+	—	(45, 38)
113	<u>Asp</u>	Asp	<u>Asp</u>	Pore, constriction zone	+	—	(37, 45)
131	Gly	<u>Lys</u>	Gln	Pore, constriction zone	+++‡	—	(47)
132	<u>Arg</u>	<u>Arg</u>	<u>Arg</u>	Pore, constriction zone	+	—	(37, 45)
164	Asp	<u>Arg</u>	Arg	L4	—	+	(47, 57)
205	Leu	<u>Arg</u>	Ile	L5	—	+	(47, 56)
248	Ser	<u>Gly</u>	Leu	L6	—	+	(56)
285	Gly	<u>Gly</u>	Arg	L7	—	+	(47, 56)

* The numbering scheme is based on the OmpF sequence (Fig. 1a, c). Residues that have been mutated and analysed are underlined. Deletion mutants are indicated by the symbol Δ. They all have increased antibiotic sensitivity. The residue range indicated for the OmpF deletion is the range covered by five different mutants.

† Effects of mutations on solute flux, phage and monoclonal antibody binding. Those with enlarged pore size were obtained by selective pressure. Altered selectivity was tested with mutants obtained by site-directed mutagenesis: Minus, no effect; ±, minor effect; +, significant effect; ++, large effect.

‡ The PhoE mutant Lys 131 → Glu shows the most pronounced change in selectivity (roughly 10-fold) and impaired polyphosphate binding⁴⁷.

the smooth end of the barrel that would contact head groups of the membrane are mainly negatively charged. The hydrophilic portion at the rough end of the barrel is larger (~20 Å high) and has a less uniform distribution of charges. Well conserved residues in this region include anionic side chains hydrogen-bonded to tyrosines in the aromatic boundary and a cluster of carboxyl groups (Fig. 2b) that could interact with the LPS through divalent cations.

Extensive studies on point mutants (Table 2) and linker insertion mutants of PhoE, which show impaired phage and antibody binding, have led to a topology model²⁰ that is in good agreement with the three-dimensional structure. The mutations are all located in loops (L2, L4–L8) thus showing that the rough end of PhoE is exposed at the cell surface. For OmpF, the results of epitope mapping predicted the same orientation³⁴, although the assignment of the position of one of the epitopes is incompatible with the structure. The spontaneous assembly of vesicles *in vitro* seems to favour an orientation opposite to that *in vivo*, because EM image reconstructions of such specimens showed that OmpF trimers lie with the smooth end outside²⁷. The proposed *in vivo* orientation agrees with the rough appearance of the surface of freeze-dried and shadowed cells³⁵. The external loops are rigid and closely packed, a finding compatible with the resistance of porins to proteases which is due to constant exposure of cells to high selection pressure.

In the three-dimensional crystals, the trimers form two-dimensional layers involving direct van der Waals interactions between hydrophobic side-chains from the barrel walls (Fig. 3). There is no room for detergent between neighbouring columns of trimers, but detergent probably covers the exposed hydrophobic surfaces near the *c*-axis of the unit cell¹⁵, in a space with a radius of 27 Å. These trigonal crystals are the first examples of well ordered type I membrane protein crystals³⁶ where the hydrophobic surface of the protein is not surrounded by a belt of detergent and where direct hydrophobic contacts exist between proteins.

Stability of the β-barrel

The closed topology of a β-barrel ensures that all of the polar main-chain atoms are engaged in interstrand hydrogen-bonds and not exposed to the membrane core. In this respect a β-barrel fulfils the same requirements as an α-helix. A protein with a missing hydrophobic core might be thought of as deformable as it apparently lacks one of the major stabilizing forces of structure as inferred from soluble proteins. Within the trimer,

however, there is a hydrophobic core up to 15 Å thick around the symmetry axis. This might explain both the stability of porin as well as the observation that subunit dissociation is accompanied by denaturation of the protein¹¹. The internal loop structure also contributes to the stability of the barrel and may affect its shape. Mutants where up to 15 residues in L3 are deleted (Table 2) are still functional³⁷, though they seem less stable³⁸. The different barrel shapes of *Rb. capsulatus* porin and the *E. coli* porins can be correlated with their differing internal loop structures. Other contributions to stability and shape of the barrel include the rigid structure formed by the closely packed, ionizable residues in the constriction zone of the pore, and a brace-forming network of tyrosine residues from the barrel walls which make hydrogen-bonds to residues in neighbouring strands and to residues from the internal loop.

Prediction of transmembrane β-strands from the amino-acid sequence is a challenging task because the hydrophobic core of the membrane can be spanned by just 7 to 9 residues with every second residue being hydrophobic (Fig. 1a). For pore-forming β-barrels, an alternation of hydrophobic and hydrophilic residues in the β-strands was postulated³⁹. This requirement is not strict, however, because internal residues can be hydrophobic if they are buried by internal structure.

The transmembrane channel

The aqueous pore has a lining similar to the outer surface of water-soluble proteins. Its framework is the 16-stranded β-barrel. Figure 4a shows that the entrance is narrowed by the long loops at the rough end of the barrel (the mouth). The internal loop L3 constricts the pore to an elliptical cross-section of 7 × 11 Å, at about half the height of the barrel. At its smooth end the pore is delineated by side chains from the barrel wall itself. Thus the pore diameter increases abruptly just beyond the constriction zone to 15 × 22 Å. Considering the channel dimensions observed, the pores are in an open form in the crystals. A molecule of glucose in its hydrated form can pass through the pore without steric hindrance⁸. Measurements of channel dimensions for OmpF, PhoE and OmpC, based on solute exclusion^{40,41} and permeability rates⁴², generally yield reliable results, though flexibility of molecules and the Stokes radii assumed for hydrated solutes introduce uncertainties. Calculations based on channel conductance^{5,6,8} are unreliable⁴³ because of the assumption that the pore is a cylinder of constant diameter over the whole width of the membrane.

The interactions between L3 and the barrel wall involve many

buried charged and polar residues. Voltage gating, which might be caused by a conformational transition involving residues in L3, could expose some of these residues, although the rigidity of L3 in the crystals does not seem to favour it. Interestingly, the sequence PEEGG (residues 116–120), which forms a turn at the tip of L3 (Fig. 4b, c) is well conserved in a porin superfamily consisting of 14 proteins⁴⁴. Pro 116 initiates the turn whereas Glu 117 is part of the pore constriction zone. Phe 118 fits into a hydrophobic pocket between L3 and the barrel wall. The Gly-Gly sequence is probably conserved because of its unique torsion angles. Deletion mutants of OmpC and OmpF isolated on the basis of growth on maltodextrins in the absence of LamB porin have larger exclusion sizes than the wild-type proteins^{37,45}. These deletions (Table 2) are all located in L3, confirming that this loop contributes significantly to the determination of the exclusion size.

The constriction zone is delimited on one side by main-chain carbonyl groups and the side chains of Asp 113 and Glu 117 from L3, and on the opposite side by a row of three arginines (Arg 42, Arg 82 and Arg 132) and Lys 16 (Fig. 4b, c). The guanidinium groups of the arginines are stacked, roughly at the same height, and presumably have favourable π - π interactions. This unusual cluster is stabilized from above and below by the side chains of Asn 64 (Glu in PhoE) and Glu 62, respectively. Tyrosines, such as Tyr 106 which points directly into the pore from L3, provide further stabilizing interactions. Owing to the clustering of charges, the residues in the constriction zone should have unusual pK values. Segregation of charged residues across the pore also exists in *Rb. capsulatus* porin²⁸. Recent theoretical calculations³⁰ of the overall electrostatic potential in this channel did not mention the ionization state of these groups. Clearly, the most significant difference in the charge distribution in *Rb. capsulatus* porin is due to the larger number of carboxyl groups from loop L3 that are involved in calcium binding. The current OmpF electron density reveals several ordered water molecules in the constriction zone, further narrowing the effective diameter of the pore.

The constriction-site residues mentioned above are strictly conserved between OmpF and PhoE, and also OmpC. But by applying strong selection pressure, mutants could be obtained^{37,45} (Table 2). The effect of single substitutions of

Arg 42, Arg 82, Asp 113 and Arg 132 by smaller residues in OmpF and OmpC strongly support the notion that the size exclusion limit of the constriction zone of the channel is determined by these residues, illustrating the power of the selection procedure. The increase in channel conductance of the OmpC mutants is commensurate with the reduction in size of the side chain concerned⁴⁶. The structural basis for the differences in selectivity, slight cation specificity of OmpF and the preference for diffusion of anions in PhoE⁵, is readily explained by the presence in PhoE of a lysine substituting for Gly 131 of OmpF that protrudes into the pore and extends the row of basic residues in the constriction zone (Fig. 4c). Site-directed mutagenesis showed that this residue, as part of a proposed anion-binding site⁴⁷, crucially influences selectivity in PhoE. Residues in the mouth of the pore, such as Lys 69 of PhoE, also have some effect on selectivity, whereas residues below the constriction zone, such as Lys 18, have minor effects (Table 2). It is interesting to note that the charge distribution at the smooth end of the barrel is similar for OmpF, PhoE and OmpC, whereas at the entrance a substantial number of differences occur.

The journey of a solute molecule through the channel may now be depicted as follows. The initial screening with respect to size and charge is done by the entrance which is restricted by the loops at the mouth of the pore. The molecule then encounters the constriction zone (~9 Å long) about halfway across the membrane, where side chains determine the size limitation and ion selectivity of the pore. After this narrow passage the molecule is effectively released into the bulk solvent. Solutes passing through the pore are, of course, hydrated, as is the pore lining. Hydrophobic molecules have a more highly structured hydration shell and therefore pass through the gate with a permeability coefficient significantly smaller than that in bulk solvent³. If the conformation of side chains and thus the water structure in the pore were altered by a trigger such as an applied potential or a change in osmotic pressure, the equilibrium between open and closed channels might be affected. As it is difficult to predict which parts of the protein are involved in channel dynamics from the crystal structures alone, site-directed mutagenesis experiments, guided by theoretical calculations, may now be exploited to elucidate this phenomenon. □

Received 11 May; accepted 22 July 1992.

- Jap, B. K. & Wallian, P. J. *Q. Rev. Biophys.* **23**, 367–403 (1990).
- Schirmer, T. & Rosenbusch, J. P. *Curr. Opin. Struct. Biol.* **1**, 539–545 (1991).
- Nikaido, H. *Molec. Microbiol.* **6**, 435–442 (1992).
- Mizuno, T., Chou, M.-Y. & Inouye, M. *J. Biol. Chem.* **258**, 6932–6940 (1983).
- Benz, R., Schmid, A. & Hancock, R. E. W. *J. Bact.* **162**, 722–727 (1985).
- Dargent, B., Hofmann, W., Pattus, F. & Rosenbusch, J. P. *EMBO J.* **5**, 773–778 (1986).
- Bühler, L. K., Kusumoto, S., Zhang, H. & Rosenbusch, J. P. *J. Biol. Chem.* **266**, 24446–24450 (1991).
- Schindler, H. & Rosenbusch, J. P. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3751–3755 (1978).
- Henderson, R. & Unwin, P. N. T. *Nature* **257**, 28–32 (1975).
- Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. *Nature* **318**, 618–624 (1985).
- Rosenbusch, J. P. *J. Biol. Chem.* **249**, 8019–8029 (1974).
- Garavito, R. M. & Rosenbusch, J. P. *J. Cell Biol.* **86**, 327–329 (1980).
- Garavito, R. M., Jenkins, J., Jansonius, J. N., Karlsson, R. & Rosenbusch, J. P. *J. molec. Biol.* **164**, 313–327 (1983).
- Weiss, M. S., Wacker, T., Weckesser, J., Welte, W. & Schulz, G. E. *FEBS Lett.* **267**, 268–272 (1990).
- Paupit, R. A. *et al. J. struct. Biol.* **107**, 136–145 (1991).
- Rossmann, M. G. *The Molecular Replacement Method* (Gordon & Breach, New York, 1972).
- Richardson, J. *Adv. Protein Chem.* **34**, 167–330 (1981).
- Lesk, A. M., Brändén, C.-I. & Chothia, C. *Protein Struct. Funct. Genet.* **5**, 139–148 (1989).
- Struyvé, M., Moons, M. & Tommassen, J. *J. molec. Biol.* **218**, 141–148 (1991).
- Bosch, D., Scholten, M., Verhagen, C. & Tommassen, J. *Molec. Gen. Genet.* **216**, 144–148 (1989).
- Rosenbusch, J. P. & Müller, R. in *Solubilization of Lipoprotein Complexes* (eds Peeters, H. & Massue, J. P.) 59–68 (European, Ghent, 1977).
- Jap, B. K., Downing, K. H. & Wallian, P. J. *J. struct. Biol.* **103**, 57–63 (1990).
- Sass, H. J. *et al. J. molec. Biol.* **209**, 171–175 (1989).
- Jap, B. K. *J. molec. Biol.* **205**, 407–419 (1989).
- Hönger, A., Gross, H., Aebi, U. & Engel, A. *J. struct. Biol.* **103**, 185–195 (1990).
- Jap, B. K., Wallian, P. J. & Gehring, K. *Nature* **350**, 167–170 (1991).
- Engel, A., Massalski, A., Schindler, H., Dorset, D. L. & Rosenbusch, J. P. *Nature* **317**, 643–645 (1985).
- Weiss, M. S. *et al. FEBS Lett.* **280**, 379–382 (1991).
- Lewis, B. A. & Engelman, D. M. *J. molec. Biol.* **166**, 211–217 (1983).
- Weiss, M. S. *et al. Science* **254**, 1627–1630 (1991).
- Deisenhofer, J. & Michel, H. *Science* **245**, 1463–1473 (1989).

- Henderson, R. *et al. J. molec. Biol.* **213**, 899–929 (1990).
- Spurlino, J. C., Lu, G.-Y. & Quiocho, F. A. *J. Biol. Chem.* **266**, 5202–5219 (1991).
- Klebba, P. E. *et al. J. Biol. Chem.* **265**, 6800–6810 (1990).
- Steven, A. C., ten Heggeler, B., Müller, R., Kistler, J. & Rosenbusch, J. P. *J. Cell Biol.* **72**, 292–301 (1977).
- Michel, H. *Trends biochem. Sci.* **8**, 56–59 (1983).
- Benson, S. A., Occi, J. L. L. & Sampson, B. A. *J. molec. Biol.* **203**, 961–970 (1988).
- Rocque, W. J. & McGroarty, E. J. *Biochemistry* **29**, 5344–5351 (1990).
- Vogel, H. & Jähnig, F. *J. molec. Biol.* **190**, 191–199 (1986).
- Hancock, R. E. W. *J. Bact.* **169**, 929–933 (1987).
- Nikaido, H. & Rosenberg, E. Y. *J. gen. Physiol.* **77**, 121–135 (1981).
- Nikaido, H. & Rosenberg, E. Y. *J. Bact.* **153**, 241–252 (1983).
- Rosenbusch, J. P. *Experientia* **46**, 167–173 (1990).
- Jeanteur, D., Lakey, J. H. & Pattus, F. *Molec. Microbiol.* **5**, 2153–2164 (1991).
- Misra, R. & Benson, S. A. *J. Bact.* **170**, 3611–3617 (1988).
- Lakey, J. H., Lea, E. J. A. & Pattus, F. *FEBS Lett.* **278**, 31–34 (1991).
- Bauer, K., Struyvé, M., Bosch, D., Benz, R. & Tommassen, J. *J. Biol. Chem.* **264**, 16393–16398 (1989).
- Messerschmidt, A. & Pfugrath, J. W. *J. appl. Crystallogr.* **20**, 306–315 (1987).
- Paupit, R. A. *et al. J. molec. Biol.* **218**, 505–507 (1991).
- Steiert, M., Ghosh, R., Schirmer, T. & Rosenbusch, J. P. *Experientia* **48**, A22 (1992).
- CCP4, SERC UK Collaborative Computing Project no. 4 (Daresbury Laboratory, Warrington, UK, 1979).
- Wang, B.-C. *Meth. Enzym.* **115**, 90–112 (1985).
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brünger, A. T. *X-PLOR Manual* (Yale Univ., New Haven, CO, 1990).
- Jones, T. A. & Kjeldgaard, M. *Manual for O* (Uppsala Univ., Uppsala, Sweden, 1990).
- Van der Ley, P., Struyvé, M. & Tommassen, J. *J. Biol. Chem.* **261**, 12222–12225 (1986).
- Korteland, J., Overbeek, N., De Graaff, P., Overduin, P. & Lugtenberg, B. *Eur. J. Biochem.* **152**, 691–697 (1985).
- Priestle, J. *J. appl. Crystallogr.* **21**, 572–576 (1988).

ACKNOWLEDGEMENTS. We thank the EMBL staff at DESY, Hamburg, for technical assistance with data collection. This study was supported by grants from the Swiss National Science Foundation to J.N.J. and J.P.R. Coordinates will be deposited in the Brookhaven Protein Databank.

Starbursts in the nuclei of shell galaxies

Lars Hernquist & Melinda L. Weil

Board of Studies in Astronomy and Astrophysics,
University of California at Santa Cruz, Santa Cruz, California 95064, USA

THE luminosity profile of elliptical galaxies for the most part follows a smooth distribution of the de Vaucouleurs form¹, but recent observations have shown that most ellipticals possess fine structure, including sharp-edged 'shells'²⁻⁵ and 'ripples'⁶⁻⁸. This suggests that the underlying dynamical structure is not fully relaxed, and Schweizer's proposal⁶ that these features result from the accretion of matter from a less massive companion galaxy has been amply confirmed by simulations of galaxy encounters and mergers⁹⁻¹⁴. A likely supply of shell-forming material is dwarf spheroidal and disk galaxies, which tend to be gas-rich, but shell galaxies do not seem to have any notable excess of gas^{15,16}. Here we report simulations of shell-forming galaxy mergers in which gas and stars are tracked separately. We find that the two components are effectively segregated: the stars form sharp-edged features by oscillating back and forth in their orbits, but most of the gas settles into a compact disk or ring in the nucleus of the primary. Shocks in the gas are likely to lead to star formation, perhaps accounting for the presence of young stars in the nuclei of some shell galaxies; if the primary galaxy contains a central black hole, evolution of the new supply of gas may yield an active galactic nucleus. Mergers between large galaxies and gas-rich satellites may therefore be a way of activating dormant elliptical galaxies.

Although recent modelling has shown, contrary to earlier expectation, that interacting galaxies need not have very different masses to produce shells^{17,18}, many examples of this kind of structure seem to have originated from the accretion of mass from a small companion galaxy onto a much larger primary. Because such companions are rich in gas, the accretion scenario predicts that there should be some correlation between the presence of shells in ellipticals and their gas content. There is limited observational evidence on this point, but the few shell ellipticals that have been examined in detail seem unremarkable in the amount of gas that they contain. Whether this is a challenge to the merger explanation for shell structure is unclear for two reasons: the observations are meagre, and nothing is known theoretically about the fate of gas accreted from a small galaxy onto a larger one. Because of recent evidence that most early-type galaxies show fine structure⁸, this is an important question to resolve.

To study this problem, we have done numerical simulations of the tidal disruption by more massive primaries of small dwarf galaxies containing both gas and stars, and the evolution of the ensuing debris. The calculations were done with a hybrid *N*-body/hydrodynamics code¹⁹. This algorithm uses a conventional *N*-body prescription to evolve collisionless matter and a particle-based technique known as smoothed particle hydrodynamics to evolve gas^{20,21}. Owing to the large mass ratio between the primary and companion, we choose to neglect the self-gravity of both galaxies, for simplicity, and use a restricted three-body method²². Following Hernquist and Quinn^{12,13}, we model the primary and companion by rigid potentials. Particles evolve in the time-varying field provided by solving the two-body problem consisting of the primary and secondary. It is assumed that the companion is instantaneously disrupted once the separation between the two galaxies is smaller than one-tenth the scale length of the primary. Thereafter, the particles evolve in the lone potential well of the primary. Here we represent the potentials of both galaxies by a simple analytical model which mimics the mass distribution in elliptical galaxies²³.

In addition to the gravitational forces which act on all particles, the gas is subject to local hydrodynamic accelerations

arising from pressure gradients and shocks. For simplicity, and owing to the rapid cooling times of interstellar gas, the calculations use an isothermal equation of state for the gas so that its temperature is always 10^4 K (ref. 24).

Figure 1 shows an example of the evolution of a rotating stellar disk accreted by a more massive spherical galaxy. In this simulation, the encounter was precisely radial and parabolic, so that the galaxies approached at escape speed, and the plane of the disk coincided with the orbit plane. Time, t , is measured in a dimensionless system of units in which the scale length and mass of the primary potential and the gravitational constant are all unity. Scaled to values appropriate for large ellipticals, a time unit corresponds to a physical time of $\sim 4 \times 10^6$ years. The disk has a mass equal to one-tenth the mass of the primary.

During its passage through the centre of the primary, which occurs between $t = 30$ and $t = 40$, the disk is destroyed by the tidal forces acting on it. The subsequent evolution of the debris is similar to that which has been reported previously⁹⁻¹⁴. Owing to the internal rotational velocities of the disk, the debris has a range of binding energies in the primary potential. As the particles oscillate back and forth through the primary, they spend more time near the turning points in their orbits, where their velocities are relatively low, leading to a projected density which is locally enhanced. As slightly less tightly bound particles turn in their orbits further out, these crests appear to propagate outwards, as waves. With time, a complex pattern of these sharp edges develops, resembling those seen in shell galaxies, although the features in this simulation are strictly two-dimensional (see refs 12, 13 for a discussion).

A rather different fate awaits the gas in this disk. Figure 2 shows the distribution of disk gas at the same times as the stars in Fig. 1. The gas in the disk makes up 10% of its mass, implying that the gas mass is 1% of that of the primary. Following the disruption of the companion, the gas particles, like the stars, begin to oscillate back and forth in the potential of the primary. With time, however, the gas distribution diverges rapidly from that of the stars. By the final time shown, most of the gas still bound to the primary has already settled into a compact spinning disk.

The differences between the evolution of stars and gas are at first sight rather puzzling, but in retrospect are straightforward to interpret. The stars, being collisionless, are free to evolve in the primary potential unimpeded by one another. The gas, however, is subject to its local self-interaction. In particular, as tightly bound gas begins to oscillate in the primary potential, it encounters less tightly bound gas near the centre moving in the opposite direction, possibly falling through the primary for the first time. These oppositely directed gas flows seed the development of strong shocks which quickly dissipate the orbital kinetic energy of the gas, leading to the formation of a dense gas clump. As less tightly bound gas falls into the centre of the primary, it collides with this clump, losing its kinetic energy. This process continues indefinitely, until most of the bound gas aggregates into a disk near the centre of the primary. The build-up of this structure is surprisingly rapid, occurring on a timescale comparable to the crossing time of the most tightly bound material.

One worry which arises in connection with the results shown in Fig. 2 concerns the special initial conditions used in this model: a radial, planar collision. But additional modelling demonstrates that the qualitative features are not altered by variations in orbital geometry. If the orbit is not precisely radial, stars and gas are again segregated: the stars form shells and, rather than settling into a nuclear disk, the gas instead forms a compact ring near the centre of the primary. A similar result obtains in nonplanar collisions, but the final structures are tilted with respect to the orbit plane.

It is perhaps not surprising that stars and gas should behave differently in situations like that depicted in Figs 1 and 2, but the efficiency of the process separating the two components is striking and could not have been predicted without detailed

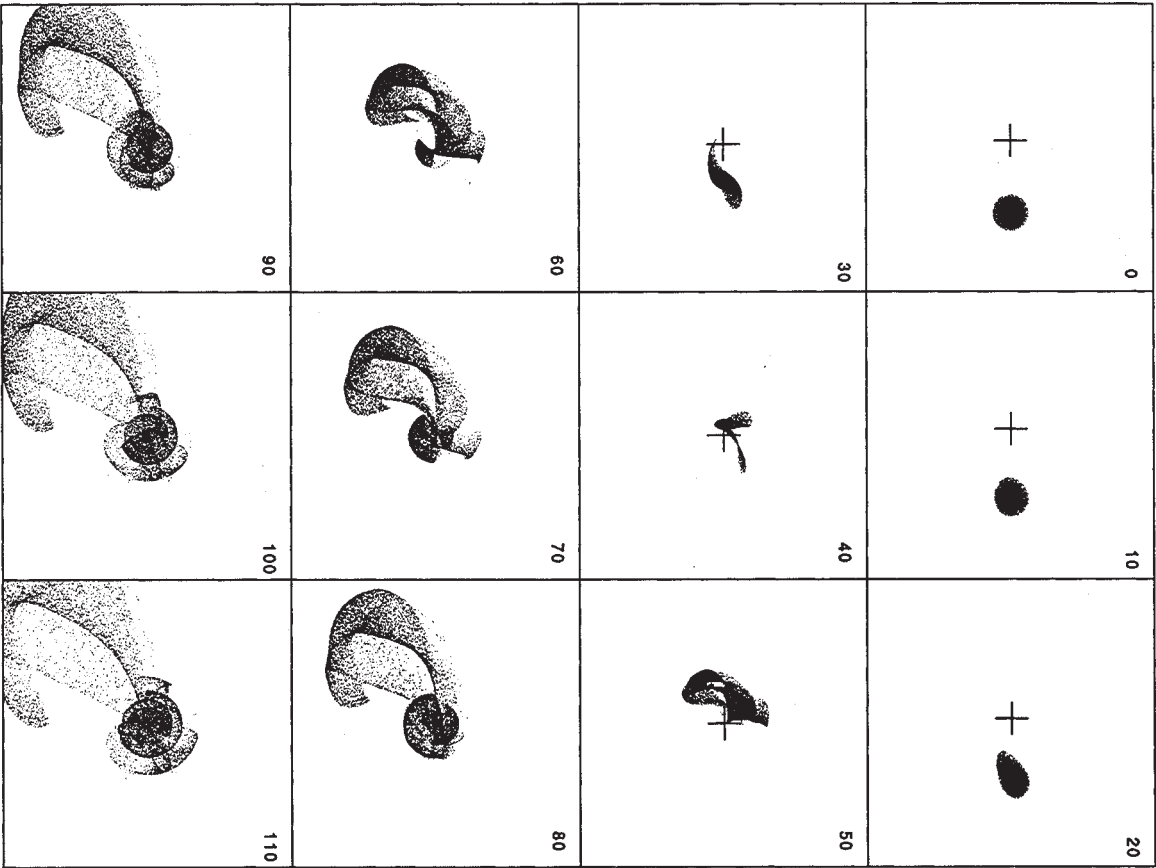


Fig. 1 Stellar component of radial encounter between a spinning disk companion of mass $m_s=0.1$ and scale length $a_s=0.7$ and a spherical primary with mass and scale length equal to unity. The initial separation of the two galaxies is 10 primary scale lengths, the galaxies approach at escape speed, and the disk is rotating counterclockwise in this projection. Dimensionless time is shown in the upper right-hand corner of all frames. All frames show the x-y spatial projection and measure 40 length units per edge. Crosses in top six panels locate centre of primary.

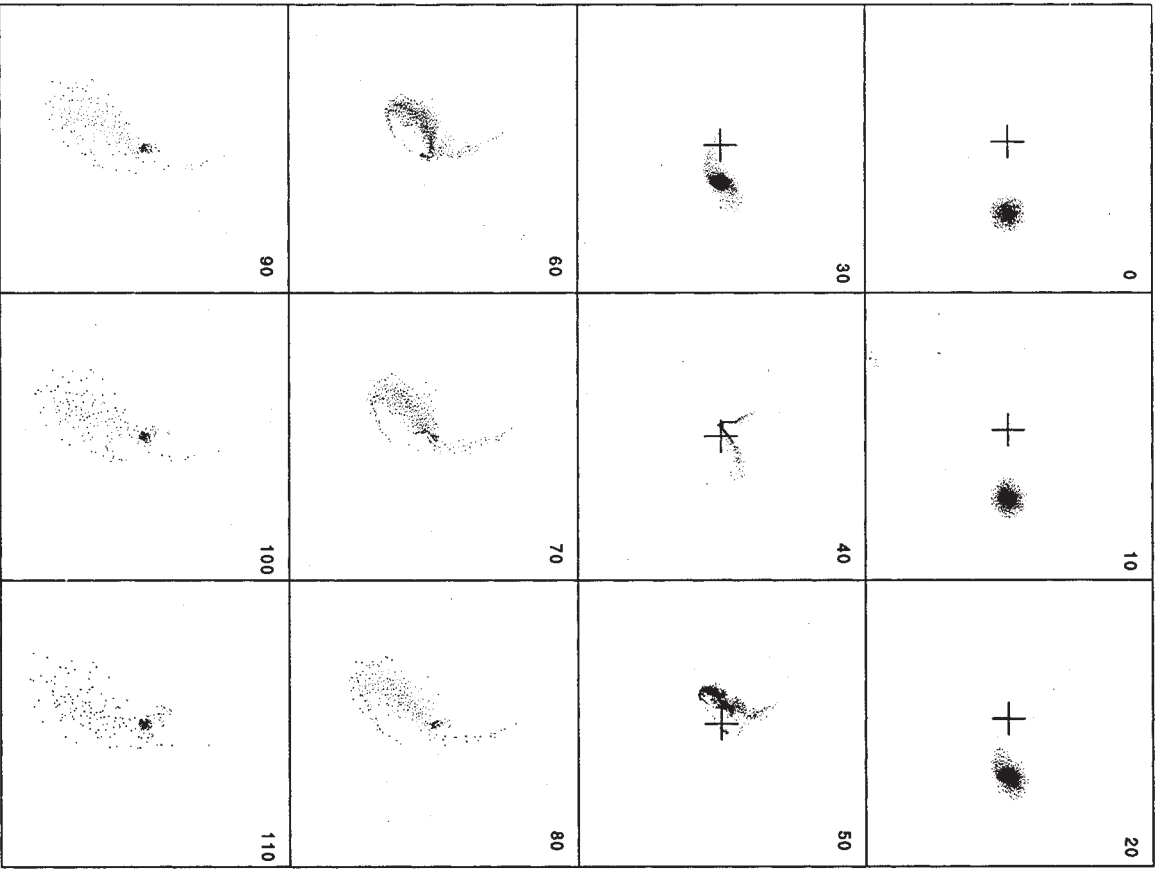


Fig. 2 Gas component of the encounter in Fig. 1. The gas disk has a mass $m_g=0.01$ and scale length $a_g=0.7$.

simulation. In particular, our results imply that gas and stars accreted by massive galaxies will be effectively 'divorced' from one another within a single crossing time, and evolve independently thereafter. This finding affects our understanding of fine structures in galaxies. Figures 1 and 2 indicate that little gas will be found near shells. Thus, shells are mainly stellar-dynamical by nature and should have only low rates of star formation. Consequently, shells should age as would an isolated stellar population and become redder with time, and we predict a weak but definite correlation between shell colours and the time elapsed since the accretion event that formed them.

A more important point is that our results severely discriminate theoretical proposals for the origin of fine structures in galaxies in that we predict that the immediate environs of shells in galaxies should be gas-poor. This is certainly not true for other proposed shell-forming mechanisms, most of which are explicitly gas-dynamical. Shells produced by shocks in an outwardly directed wind, by blast waves or by accretion of intergalactic gas²⁵⁻²⁸ require an abundant supply of fresh gas and ought to be gas-rich. These considerations should motivate high-sensitivity searches for the presence of gas in galaxies containing shells. For example, the discovery of shells rich in gas would contradict our simulation results and argue against their having originated by accretion events. As fine structures are ubiquitous, such a finding would profoundly alter our understanding of the formation and evolution of galaxies.

The ultimate fate of the gas in the models is problematic, as the effects of star formation and associated energy input are ignored. Nevertheless, it seems plausible that star formation will be boosted in the nuclear disks and rings formed during accretion events by the strong shocks responsible for dissipating the orbital kinetic energy of the gas. This suggests that a distinct population of stars will be produced near the centre of the primary. A considerable fraction of shell galaxies (~20%) in the Malin-Carter catalogue⁵ contain nuclear concentrations of A stars, indicative of recent star formation²⁹. The shell galaxy 0140-658 possesses a post-starburst spectrum which is quite blue in the nucleus and has a steep gradient. The galaxy Markarian 717 displays a starburst nucleus and shells³⁰. It is natural to suggest that the activity in these objects was triggered by an accretion event like that depicted in Figs 1 and 2 which simultaneously produced the shells around the galaxies.

More speculatively, our results may be relevant to the interpretation of recent observations of NGC1275, an elliptical galaxy which possesses shells and a central population of bright, blue point sources with characteristics similar to those of young globular clusters³¹. These sources lie within 5 kpc of the centre of the elliptical; their distribution resembles that of the gas accreted in the models shown here.

Finally, it is recognized that a small fraction of ellipticals contain active galactic nuclei. Popular consensus suggests that the emission from these objects arises from gas accretion onto a supermassive black hole. The supply of gas that might power black holes in ellipticals is uncertain, as these galaxies are intrinsically gas-poor. Although the dynamic range of our models is not large enough for us to predict whether the central gas accumulations can power activity on very small scales, several phenomenological arguments suggest that nuclear disks and rings should be dynamically unstable, leading to inflow beyond that seen here³². If future modelling verifies these conjectures, then our simulations provide a plausible means of triggering activity in the nuclei of elliptical galaxies. □

7. Schweizer, F. in *Dynamics and Interactions of Galaxies* (ed. Wielen, R.) 60-71 (Springer, Berlin, 1990).
8. Seitzer, P. & Schweizer, F. in *Dynamics and Interactions of Galaxies* (ed. Wielen, R.) 270-271 (Springer, Berlin 1990).
9. Quinn, P. J. thesis, Australian National Univ. (1982).
10. Quinn, P. J. *Astrophys. J.* **279**, 596-609 (1984).
11. Toomre, A. in *IAU Symp. 100: Internal Kinematics and Dynamics of Galaxies* (ed. Athanassoula, E.) 319 (Reidel, Dordrecht, 1983).
12. Hernquist, L. & Quinn, P. J. *Astrophys. J.* **331**, 682-698 (1988).
13. Hernquist, L. & Quinn, P. J. *Astrophys. J.* **342**, 1-16 (1989).
14. Heisler, J. & White, S. D. M. *Astrophys. J.* **243**, 199-208 (1990).
15. Wilkinson, A. et al. *Mon. Not. R. astr. Soc.* **224**, 895-910 (1987).
16. Wilkinson, A., Browne, I. W. A. & Wolstencroft, R. D. *Mon. Not. R. astr. Soc.* **228**, 933-940 (1987).
17. Barnes, J. *Astrophys. J.* (in the press).
18. Hernquist, L. *Astrophys. J.* (in the press).
19. Hernquist, L. & Katz, N. *Astrophys. J. Suppl.* **70**, 419-446 (1989).
20. Gingold, R. A. & Monaghan, J. J. *Mon. Not. R. astr. Soc.* **181**, 375-389 (1977).
21. Lucy, L. *Astr. J.* **82**, 1013-1024 (1977).
22. Toomre, A. & Toomre, J. *Astrophys. J.* **178**, 623-666 (1972).
23. Hernquist, L. *Astrophys. J.* **356**, 359-364 (1990).
24. Hernquist, L. & Barnes, J. E. *Nature* **354**, 210 (1991).
25. Bertschinger, E. *Astrophys. J. Suppl.* **58**, 39-72 (1985).
26. Fabian, A. C., Nulsen, P. E. J. & Stewart, G. C. *Nature* **287**, 613-614 (1980).
27. Löwenstein, M., Fabian, A. C. & Nulsen, P. E. J. *Mon. Not. R. astr. Soc.* **229**, 129-141 (1987).
28. Williams, R. E. & Christiansen, W. A. *Astrophys. J.* **291**, 80-87 (1986).
29. Carter, D., Prieur, J. L., Wilkinson, A., Sparks, W. B. & Malin, D. F. *Mon. Not. R. astr. Soc.* **235**, 813-825 (1988).
30. Taniguchi, Y., Ichikawa, S., Hamabe, M., Yamagata, T. & Iye, M. *Astr. Astrophys.* **233**, 385-390 (1990).
31. Holtzmann, J. A. et al. *Lowell Observatory Preprint; Astr. J.* **103**, 691 (1992).
32. Gunn, J. E. in *Active Galactic Nuclei* (ed. Hazard, C. & Mitton, S.) 213-225 (Cambridge Univ. Press, 1979).

ACKNOWLEDGEMENTS. We thank F. Schweizer for discussion. This work was supported in part by an allocation of Cray time from the Pittsburgh Supercomputing Center, the Alfred P. Sloan Foundation, NASA, the California Space Institute, the NSF and the Presidential Faculty Fellows Program.

Acceleration of the rate of nitrite oxidation by freezing in aqueous solution

Norimichi Takenaka, Akihiro Ueda & Yasuaki Maeda

Environmental Chemistry, College of Engineering,
University of Osaka Prefecture, 1-1 Gakuen-cho, Sakai 593, Japan

WHEN a dilute ionic solution freezes, differences in the partitioning of ions in the aqueous and ice phases can generate electric potentials which may influence electrochemical reactions¹⁻³. Pitter and co-workers^{4,5} have studied the influence of freezing on the endothermic oxidation of sulphide to sulphate in growing ice crystals inside a cloud chamber. Here we report that the oxidation of nitrite by dissolved oxygen to form nitrate, which is a very slow process in solution, is accelerated markedly when it takes place in a solution undergoing freezing. At pH 4.5 and a temperature of 25 °C, the rate is increased by a factor of about 10⁵ for a freezing rate of 0.2 g solution per minute; the reaction rate increases as the freezing rate increases. Although the mechanism of this acceleration is not yet clear, we are able to eliminate the possibilities of thermochemical, photochemical and simple electrochemical reactions, and of catalysis on the ice surface. Processes of this sort may be important for chemical reactions taking place in freezing cloud and fog droplets in the atmosphere.

Reagent-grade chemicals were obtained from Wako Pure Chemicals Inc. and used without further purification. NaNO₂ was dried at 120 °C for 10 hours. We prepared an aqueous solution by dissolving NaNO₂ in Milli-Q water (resistivity >18 MΩ cm) and bubbled an O₂/N₂ mixture through the solution for more than 30 min to adjust the concentration of dissolved oxygen. The pH of the solution was adjusted with sulphuric or formic acid just before freezing. We measured the concentrations of anions, Na⁺, H⁺ and dissolved oxygen by ion chromatography, flame photometry, a pH meter and gas chromatography with electron-capture detection⁶, respectively. Species frozen into the ice were analysed after thawing. To measure ice formation rates, we separated the ice and solution completely by suction of the solution and then measured their weights.

Received 18 March; accepted 17 July 1992.

1. de Vaucouleurs, G. *Ann. Astrophys.* **11**, 247-287 (1948).
2. Arp, H. *Astrophys. J. Suppl.* **14**, 1-20 (1966).
3. Malin, D. F. *Nature* **277**, 279-280 (1979).
4. Malin, D. F. & Carter, D. *Nature* **285**, 643-645 (1980).
5. Malin, D. F. & Carter, D. *Astrophys. J.* **274**, 534-540 (1983).
6. Schweizer, F. *Astrophys. J.* **237**, 303-318 (1980).

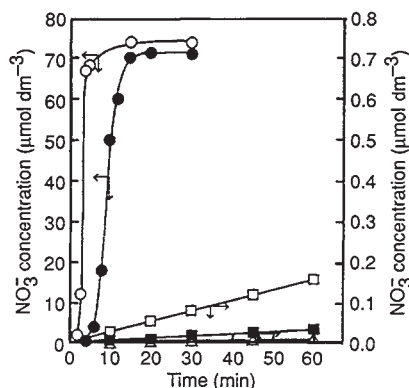


FIG. 1 Time profile of the reaction of HNO_2 with dissolved oxygen at pH 4.0. $\circ$, Freezing process with coolant at -45°C ; $\bullet$, freezing process with coolant at -21°C ; $\square$, in solution at 60°C ; $\blacksquare$, in solution at 25°C ; $\triangle$, in solution at 4°C . Different scales are used for results on freezing and in solution. The initial concentration of N(III) was $1.0 \times 10^{-4} \text{ mol dm}^{-3}$, and that of dissolved oxygen was 4.4×10^{-4} for the freezing process and $2.5 \times 10^{-4} \text{ mol dm}^{-3}$ for the solution.

Each sample solution was frozen in the following way. An aliquot of the solution was drawn into a polyethylene syringe (75 mm long, internal diameter 15 mm) sealed with a septum (free from gas phase) to hold constant the concentration of dissolved oxygen, and the syringe was then immersed in the coolant. The solution froze from the wall to the centre. Analytical results were similar (under the same conditions) for both the closed system described above and an open system in which the sample was equilibrated with the ambient air. The results were also similar when the solution froze (at the same rate) from the bottom upwards rather than from wall to centre. The material used for the reaction vessel (glass, stainless steel or polyethylene) did not affect the reaction.

Figure 1 shows the time profile of the concentration of NO_3^- formed by reaction of N(III) with dissolved oxygen. The reaction rates in solution decreased with decreasing temperature. In contrast, the reaction rates produced by freezing were much faster than those in solution. Furthermore, the lower the temperature of the coolant (that is, the faster the freezing rate), the faster was the reaction rate.

The reaction rate produced by freezing was not affected by light in the laboratory, so this is not a photochemical reaction. The acceleration was never observed when pure crushed ice was

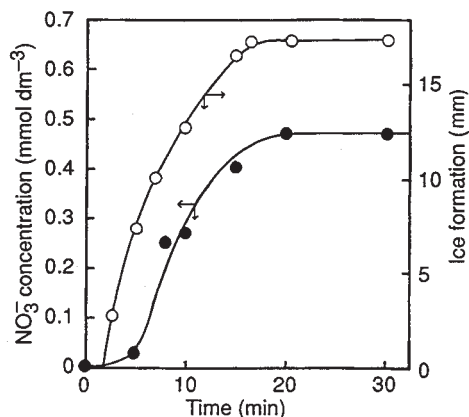


FIG. 2 Time profile of the formations of NO_3^- and ice when freezing from the bottom upwards. $\bullet$, NO_3^- formation, $\circ$, ice formation, calculated from the weight measured after separation of ice and solution. The values are height of the ice. The freezing rate was 0.2 g min^{-1} . The initial concentrations of N(III) and dissolved oxygen were 5.0×10^{-4} and $4.42 \times 10^{-4} \text{ mol dm}^{-3}$, respectively. The sample mass was 3 g.

added to N(III) solution at pH 3.0 and 0°C , so it is not a catalytic reaction on the ice surface. Furthermore, the reaction rates did not depend on the thawing rates or methods, but on the freezing rate. The reaction took place only in acid solution. Sulphuric acid and sulphate did not participate in the reaction: NO_3^- formed when we added HCOOH to adjust the pH, but it did not form when we added $1 \times 10^{-3} \text{ mol dm}^{-3}$ of SO_4^{2-} at pH 5.6 and froze the solution with $0.442 \times 10^{-3} \text{ mol dm}^{-3}$ of dissolved oxygen.

Freezing can be accompanied by the generation of an electric potential difference between ice and solution¹⁻⁵. In our results, the concentration of dissolved oxygen, which is not related to the generation of the potential difference, was observed to be correlated with the reaction rate. Figure 2 shows time profiles of ice and NO_3^- formation. The formation of NO_3^- was independent of the ice formation. Workman and Reynolds¹ reported that the potential difference was first generated when freezing started, and that it disappeared when freezing stopped. We therefore consider that the reaction is not directly related to the potential difference, and therefore is not directly accelerated by an electric effect. Finnegan *et al.*⁵ pointed out that the reaction in the growing ice crystal took place by separation of ions, resulting in the electric effect. We analysed all ions, HNO_2 and O_2 in the solution and the ice after separation, and observed that they were slightly rejected from the ice. Therefore in the present system the reaction by freezing may be also accelerated by the separation of ions.

The rate of NO_3^- formation by freezing was experimentally determined to be $0.05\text{--}1 \times 10^{-3} \text{ mol dm}^{-3}$ of N(III) , $0\text{--}0.97 \times 10^{-3} \text{ mol dm}^{-3}$ of dissolved oxygen, and pH 3.0–4.5 (the pH range observed in rain in industrial areas) under pseudo-first-order conditions. Figure 3 shows the rates of NO_3^- formation against initial concentrations of N(III) and dissolved oxygen. The rate was defined as the maximum rate in the time profile of the reaction. For N(III) and dissolved oxygen, the correlation coefficients were 0.966 and 0.978, respectively, and the slopes were 0.82 and 0.99, respectively, indicating first-order dependence for both N(III) and dissolved oxygen. From these results we obtained the following equation

$$\frac{d[\text{NO}_3^-]}{dt} = 0.4[\text{N(III)}]_0[\text{O}_2]_0 \quad (\text{mol dm}^{-3} \text{ s}^{-1}) \quad (1)$$

Here the subscript zero denotes the initial concentration. Formation of NO_3^- was negligible in the absence of dissolved oxygen. As the mass balance of N (including only nitrate and nitrite) was reasonable, N(III) must be reacting with dissolved oxygen.

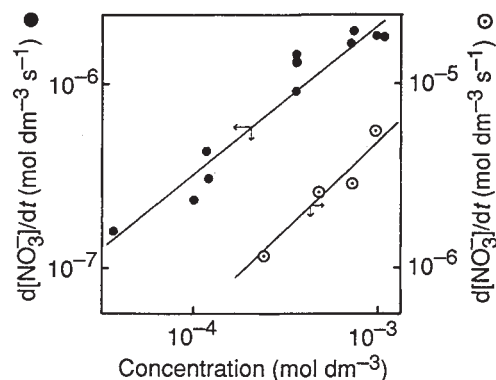


FIG. 3 The NO_3^- formation rate against the initial concentrations of N(III) and dissolved oxygen. $\bullet$, Rate against N(III) concentration. The initial concentration of dissolved oxygen and the pH were $1.9 \times 10^{-3} \text{ mol dm}^{-3}$ and 4.0, respectively. $\circ$, Rate against concentration of dissolved oxygen. The initial concentration of N(III) and the pH were $1.0 \times 10^{-2} \text{ mol dm}^{-3}$ and 4.0, respectively. The solution was frozen at 0.2 g min^{-1} .

TABLE 1 The rate of NO₃⁻ formation

Concentration (mol dm ⁻³)			Rate of NO ₃ ⁻ formation (mol dm ⁻³ s ⁻¹)		
N(III)*	Oxidant†	pH	R _s ‡	R _f §	R _f /R _s
5.28 × 10 ⁻⁶	O ₂ , 0.25 × 10 ⁻³	4.5	3.56 × 10 ⁻¹⁵		
5.28 × 10 ⁻⁶	O ₂ , 0.44 × 10 ⁻³	4.5		9.29 × 10 ⁻¹⁰	260,000
1.44 × 10 ⁻⁶	O ₂ , 0.25 × 10 ⁻³	4.0	1.90 × 10 ⁻¹⁵		
1.44 × 10 ⁻⁶	O ₂ , 0.44 × 10 ⁻³	4.0		2.53 × 10 ⁻¹⁰	133,000
1.44 × 10 ⁻⁶	O ₃ , 6.5 × 10 ⁻¹⁰	4.0	3.6 × 10 ⁻¹⁰		
1.44 × 10 ⁻⁶	H ₂ O ₂ , 2.94 × 10 ⁻⁵	4.0	4.39 × 10 ⁻¹²		

* Values of N(III) were those measured in rain in Osaka, Japan on 15–16 February and 5 March 1992.

† Values of dissolved oxygen were the concentrations equilibrated with air at 25 °C and 0 °C for the solution and the freezing process, respectively. The value of O₃ was that equilibrated with 50 parts per 10⁹ of O₃ in air⁹. The concentration of H₂O₂ was generally observed in rain in industrial area.

‡ Rates in solution were calculated according to equation (2). The rates for ozone and H₂O₂ were calculated by the method in ref. 14.

§ Rates produced by freezing were calculated according to equation (1).

The reacting species of N(III) and the reaction mechanism have not yet been clarified.

The kinetics in solution have been reported as⁷

$$\frac{d[\text{NO}_3^-]}{dt} = 72[\text{HNO}_2]^2[\text{O}_2] \quad (\text{mol dm}^{-3} \text{ s}^{-1}) \quad (2)$$

The concentration of HNO₂ was calculated by using a dissociation constant $5.9 \times 10^{-4} \exp[-1,760(1/T - 1/298)] \text{ mol dm}^{-3}$ (ref. 8). Table 1 shows the calculated rates of NO₃⁻ formation. The rates produced by freezing were more than 10⁵ times faster than those in solution. Because the reaction rates increased with increasing freezing rate, the rate of NO₃⁻ formation may exceed the above values when supercooled water droplets freeze. We also compared the reaction rate of N(III) with dissolved oxygen by freezing with the reaction rates of N(III) with O₃ and H₂O₂, the main oxidants in rain. As shown in Table 1, the reaction produced during freezing was much faster than that with H₂O₂ in solution but comparable to that with ozone. We believe that the reaction with dissolved oxygen by freezing is one of the fastest reactions that occurs in solution.

Nitrates are one of the primary constituents of acid rain, and are thought to be produced by dissolution of nitrate aerosols and gaseous HNO₃ and N₂O₅ in water droplets⁹. The mechanism has not, however, yet been clarified completely^{10,11}. The concentration of nitrite (HNO₂ and NO₂⁻) in the atmosphere is a few parts per 10⁹ in the gas phase^{12–14} and $1\text{--}5 \times 10^{-4} \text{ mol dm}^{-3}$ in rain¹⁵. Nitrite has not previously been recognized as a precursor to nitrate in atmospheric water droplets¹¹, but we suggest that if nitrite-containing droplets freeze, the nitrite could be converted in significant amounts to nitrate and nitric acid by the process reported here. Our results also support the conclusion of Pitter *et al.*⁴ and Finnegan *et al.*⁵ that freezing substances in order to preserve them might actually alter their compositions considerably. □

Received 21 April; accepted 16 July 1992.

1. Workman, E. J. & Reynolds, S. E. *Phys. Rev.* **78**, 254–259 (1950).
2. Lodge, J. P., Baker, M. L. & Pierrard, J. M. *J. chem. Phys.* **24**, 716–719 (1956).
3. Cobb, A. W. & Gross, G. W. *J. electrochem. Soc.* **116**, 796–804 (1969).
4. Pitter, R. L., Finnegan, W. G. & Young, L. G. *Proc. Int. Conf. Global and Regional Environmental Atmospheric Chemistry* (eds Newman, L., Wang, W. & Kiang, C. S.) 466–469 (U.S. Department of Energy, Washington DC, 1989).
5. Finnegan, W. G., Pitter, R. L. & Young, L. G. *Atmos. Environ.* **A25**, 2531–2534 (1991).
6. Maeda, Y. & Munemori, M. *Chem. Express* **1**, 467–470 (1986).
7. Damschen, D. E. & Martin, L. B. *Atmos. Environ.* **17**, 2005–2011 (1983).
8. Nair, S. K. & Peters, L. K. *Atmos. Environ.* **23**, 1399–1423 (1989).
9. Schwartz, S. E. *SO₂, NO and NO₂ Oxidation Mechanisms: Atmospheric Considerations* (ed. Calvert, J. G.) 173–208 (Ann Arbor Science, Boston, 1984).
10. Hegg, D. A. & Hobbs, P. V. *J. Atmos. Chem.* **7**, 325–333 (1988).
11. Hara, H. *J. Japan Soc. Air Pollut.* **25**, 1–29 (1990).
12. Platt, U., Perner, D., Harris, G. W., Winer, A. M. & Pitts, J. N. *J. Nature* **285**, 312–314 (1980).
13. Harris, G. W. *et al. Environ. Sci. Technol.* **16**, 414–419 (1982).
14. Sjodin, A. *Environ. Sci. Technol.* **22**, 1086–1089 (1988).
15. Ruprecht, H. & Sigg, L. *Atmos. Environ.* **A24**, 573–584 (1990).

ACKNOWLEDGEMENTS. We thank T. Dohmaru for discussion. This work was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture, Japan.

Thermal skin effect of the surface ocean and its implications for CO₂ uptake

J. E. Robertson*† & A. J. Watson*

* Plymouth Marine Laboratory, Prospect Place, West Hoe, Plymouth, PL1 3DH, UK

† School of Ocean Sciences, UCNW, Menai Bridge, Gwynedd LL59 5EY, UK

AN understanding of the natural sources and sinks of atmospheric carbon dioxide is necessary for predictions of future atmospheric loading and its consequences for global climate. Present estimates of emissions and uptake do not balance¹, and although some have attributed the imbalance to a terrestrial sink², the magnitude of the oceanic sink remains unresolved^{2–4}. It is known^{5–8} that the upper 1 mm or so of the oceans represents a cool 'skin', with a temperature gradient such that the surface is generally cooler than the bulk mixed layer by about 0.3 °C as a result of the upward heat flux. Here we discuss the consequences of the 'skin effect' for the global air–sea flux of CO₂. We show that it produces an increased oceanic global uptake of about 0.7 Gt C yr⁻¹, when the flux from the atmosphere to the oceans is calculated on the basis of measured surface-ocean partial pressures of CO₂. This correction helps to bring into closer agreement the oceanic CO₂ uptake calculated by different methods^{2–4}.

The most consistent estimates of global ocean uptake have been calculated through the use of oceanic models calibrated by ¹⁴CO₂ and other tracers, and lie in the range 2.2 ± 0.5 Gt carbon per year^{3,4}. In a more direct approach^{2,9}, data for surface CO₂ are used to integrate air–sea exchange over the global ocean. The flux (*F*) at any point is given by

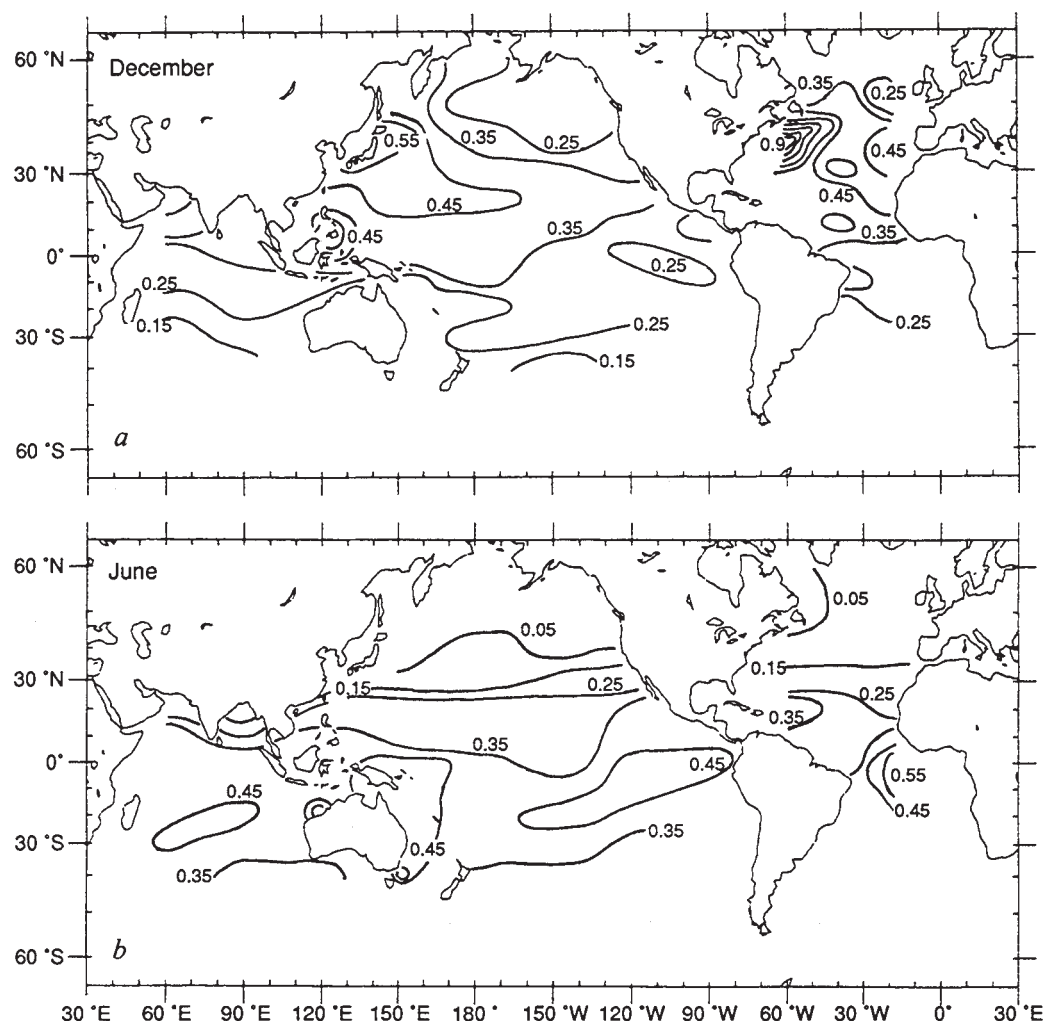
$$F = K\{[\text{CO}_{2s}] - [\text{CO}_2^*]\} \quad (1)$$

where *K* is the gas transfer velocity and depends on local wind speed and temperature, and [CO_{2s}], [CO₂^{*}] are the concentrations of dissolved gaseous CO₂ at, respectively, the surface and the base of the diffusion-dominated sublayer (~50 μm thick) which limits exchange at the surface. Normally, these concentrations are approximated by [CO_{2s}] = α_{ml}pCO_{2atm} and [CO₂^{*}] = α_{ml}pCO_{2ml}. Here α_{ml}(*T*, *S*) is the solubility as a function of temperature and salinity, calculated for conditions in the bulk mixed layer, and pCO_{2atm}, pCO_{2ml} are the partial pressures of CO₂ in the atmosphere and in equilibrium with water from the mixed layer. These substitutions give

$$F = K\alpha_{ml}[(p\text{CO}_{2\text{atm}}) - (p\text{CO}_{2\text{ml}})] \quad (2)$$

which is the equation normally applied. But because the temperature at the surface is usually lower than that in the mixed

FIG. 1 Surface skin-temperature deviations (in K) for (a) December and (b) June calculated from monthly mean climatological data and the expression proposed by Hasse⁵. The actual deviation at any point varies diurnally, being sensitive to the local solar flux. The estimates shown here, calculated with monthly mean solar fluxes, should represent unbiased averages over this diurnal cycle.



layer, the solubility is slightly higher. If we write

$$\alpha_{\text{surface}} = \alpha_{\text{ml}}(1 + \delta) \quad (3)$$

the corrected equation (2) becomes

$$F = K\alpha_{\text{ml}}[(p\text{CO}_{2\text{atm}}) - (p\text{CO}_{2\text{ml}})] + [\delta\alpha_{\text{ml}}K \cdot p\text{CO}_{2\text{atm}}] \quad (4)$$

There is now an additional flux due to the cooler surface skin.

With the present limited database of surface $p\text{CO}_2$ measurements, this approach yields lower global fluxes than the tracer-calibrated models; they lie in the range $0.3\text{--}1.3\text{ Gt C yr}^{-1}$, depending on the assumptions made about how $p\text{CO}_{2\text{ml}}$ and K vary. Although this method of calculating global fluxes is subject to errors because of the space and time variability of K and $p\text{CO}_{2\text{ml}}$ (refs 10–13), it is valuable in that it gives an independent check on the global net uptake deducted from models, and furthermore yields a detailed description of the location and magnitude of oceanic source and sink regions, which is necessary to validate 'inverse' models of the distribution of atmospheric CO_2 (refs 2, 14, 15).

Global calculations based on equation (1) are sensitive to systematic errors in the values of $p\text{CO}_{2\text{atm}}$ or $p\text{CO}_{2\text{ml}}$. For example, in the sensitivity calculations published by Tans *et al.*², systematic error of $1\text{ }\mu\text{atm}$ in the annual average estimate of $p\text{CO}_2$ in the surface water (only $\sim 0.3\%$ of the absolute value) would produce a total flux error of $\sim 0.22\text{ Gt yr}^{-1}$ carbon over the global ocean. Neglect of the thermal skin effect results in a similar error, as the solubility (α) of the gas is a strong function of temperature.

To adjust equation (1) for a skin temperature deviation ΔT from the mixed-layer temperature, $[\text{CO}_{2\text{s}}]$ must be calculated

at the true surface temperature. As $\alpha(T - \Delta T, S) \cdot p\text{CO}_{2\text{atm}} \cdot \alpha$ in sea water increases by 2.5–3.8% per degree of cooling over the range of ocean surface temperatures¹⁶, the adjustment is typically equivalent to $+2.5\text{--}4\text{ }\mu\text{atm}$ for $\Delta T = -0.3\text{ }^\circ\text{C}$ and $p\text{CO}_{2\text{atm}} = 350\text{ }\mu\text{atm}$. The second term in the equation needs no adjustments, as $[\text{CO}_2^*] = [\text{CO}_{2\text{ml}}]$ provided that transport from the mixed layer to the base of the diffusion sublayer occurs in a time short compared with that for hydration of CO_2 molecules, roughly 50 s (ref. 17).

The existence of a 'cool skin' on the top of the ocean is well known^{5–8}. The deviation from the bulk mixed-layer temperature is not universally negative⁸, but on average is a few tenths of a degree colder than the bulk temperature. The temperature deviation is maintained by an outgoing thermal flux, which generates a considerable temperature gradient in the layer where turbulence is suppressed towards the interface, the thermal flux balancing the absorption of solar radiation which takes place in the top $\sim 75\text{ m}$ of the water column. The thermal boundary layer occupies the uppermost $\sim 50\text{--}500\text{ }\mu\text{m}$ depending on local heat flux and wind stress⁶, and persists at wind speeds up to at least 10 m s^{-1} . It can re-establish itself in 10–12 s if the wind speed drops to below 10 m s^{-1} (ref. 6).

Figure 1 shows global surface ΔT due to the skin effect in June and December, calculated from monthly mean wind speed, solar radiation and heat flux with the parameterization of Hasse⁵. His expressions for the difference between surface and mixed layer temperature measured at 2.5 m depth (a typical depth for 'surface' CO_2 analyses) are

$$\Delta T = C_1 H / U + C_2 Q / U \quad (5)$$

where H is heat flux times $0.00143\text{ (W m}^{-2}\text{)}$, U is wind speed

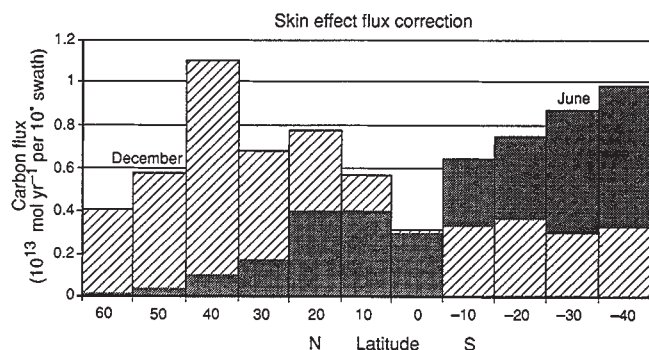


FIG. 2 Zonally averaged corrections to the flux of CO_2 from air to sea (in units of $10^{13} \text{ mol C yr}^{-1}$) implied by the surface temperature deviations shown in Fig. 1a and b. Each point is the flux correction calculated over the 10° wide zonal belt centred around that latitude. The gas transfer coefficient used was the 'empirical' relation of Tans *et al.*² calculated from the monthly mean wind.

(m s^{-1}), Q is solar radiation times $0.00143 \text{ (W m}^{-2}\text{)}$, $C_1 = 10.3$, $C_2 = 2.14$. Here fluxes are defined as positive when entering the water and negative on leaving, so Q (solar radiation) is positive and ΔT usually negative. The parameterization is derived from short-term observations and errors may occur when it is applied to monthly mean meteorological data sets. Although our calculations show that the surface skin effect is important in gas flux calculations, they would benefit from higher-resolution data.

We took heat flux and shortwave radiation data from ref. 18 and wind speed data from ref. 19. These sources include little data for the Southern Ocean south of 40° S . A notable anomaly in ΔT is seen in the North Atlantic in winter, because of the strong heat flux associated with the gulf stream and North Atlantic drift. The Northern Hemisphere oceans, particularly the North Atlantic, show a slightly larger effect than do those in the Southern Hemisphere because of the net interhemispheric heat transport from south to north.

Figure 2 shows zonally averaged corrections to the ocean-atmosphere CO_2 flux for June and December calculated using the ΔT values of Fig. 1 and equation (1) where the variation of gas transfer K with wind has been approximated by the 'empirical' relation of ref. 2 applied to the monthly climatological mean wind. The correction is positive (flux into the sea is increased) at all latitudes and is greater for the winter hemisphere than for the summer. The sum of the global flux corrections, excluding the Southern Ocean south of 45° S , is $0.56 \text{ Gt C yr}^{-1}$ in June and $0.69 \text{ Gt C yr}^{-1}$ in December. This suggests an annual average of $\sim 0.63 \text{ Gt C yr}^{-1}$, of which $0.36 \text{ Gt C yr}^{-1}$ is in the Northern Hemisphere oceans. If we conservatively estimate that the Southern Ocean south of 45° S would add 10% to the correction, the global total is $\sim 0.7 \text{ Gt C yr}^{-1}$. Although this is only $\sim 0.7\%$ of the gross sea-air flux, it is a large fraction ($\sim 30\%$) of the net uptake calculated from ocean models.

To balance the atmospheric CO_2 budget, the global net ocean-land surface sink must be $3.5 \pm 1.4 \text{ Gt C yr}^{-1}$ where most of the uncertainty is due to inadequate knowledge of the rate

of tropical deforestation¹. In addition, from inversion studies of the atmospheric north-south gradient², the Northern Hemisphere sink is constrained to be $\sim 2 \text{ Gt C yr}^{-1}$ greater than that for the Southern Hemisphere, an observation which has led to the hypothesis that terrestrial temperate vegetation (almost all in the north) is acting at present as a substantial sink². The skin effect correction is about equal for the two hemispheres and therefore cannot explain the asymmetric distribution of the global sinks. It does, however affect the relative magnitudes of terrestrial and oceanic uptake and brings estimates of these sinks derived by different techniques closer to agreement. For example, Tans *et al.*² used the calculated ocean uptake for the area north of 15° S to constrain the global budget. For this area, we calculate an average skin effect correction (mean of June and December) of $0.44 \text{ Gt C yr}^{-1}$, and this should be added to the maximum value deduced by Tans *et al.*² for uptake by the oceans of 0.8 Gt C yr^{-1} . This adjustment accounts for $\sim 50\%$ of the difference between their maximum uptake and the lower limit of 1.7 Gt C yr^{-1} calculated from tracer-calibrated ocean uptake models, and thus for 50% of the minimum discrepancy. The same amount should be subtracted from the terrestrial sink that they inferred ($2\text{--}3.4 \text{ Gt C yr}^{-1}$ depending on the poorly known source-strengths of deforestation and other possible land sources).

Two other corrections tend to reduce the discrepancy between the two methods of calculating ocean uptake. These are concerned with carbon monoxide in the atmosphere²⁰ (magnitude $\sim 0.25 \text{ Gt C yr}^{-1}$) and the riverine flux of carbon from land to sea⁴ ($0.4\text{--}0.7 \text{ Gt C yr}^{-1}$). Sarmiento and Sundquist⁴ have shown that the corrections (including a first estimate of skin temperature effect communicated by one of us) combine to allow the ocean model uptake and global air-sea flux estimates to overlap within the present uncertainties. Further refinement of the ocean source and sink regions of atmospheric CO_2 should be achievable by improving the coverage of surface ocean $p\text{CO}_2$ observations and a careful re-examination of possible systematic errors in the gas exchange calculation²¹. □

Received 31 March; accepted 14 July 1992.

- Houghton, J. T., Jenkins, G. J. & Ephraums, J. J. (eds) *Climate Change, The IPCC Scientific Assessment* (Cambridge Univ. Press 1990).
- Tans, P. P., Fung, I. Y. & Takahashi, T. *Science* **247**, 1431–1438 (1990).
- Sarmiento, J. L., Siegenthaler, U. & Orr, J. C. *J. geophys. Res.* (in the press).
- Sarmiento, J. L. & Sundquist, E. *Nature* **356**, 589–593 (1992).
- Hasse, L. *Boundary Layer Met.* **1**, 368–379 (1971).
- Katsaros, K. B. *Tellus* **29**, 229–239 (1977).
- Simpson, J. J. & Paulson, C. A. *J. phys. Oceanogr.* **10**, 399–410 (1980).
- Schluessel, P., Emery, W. J., Grasi, H. & Mammen, T. *J. geophys. Res.* **95**, 13341–13356 (1990).
- Etcheto, J., Boutin, J. & Merlivat, L. *Tellus* **43B**, 247–255 (1991).
- Watson, A. J., Robinson, C., Robertson, J. E., Williams, P. J. LeB. & Fasham, M. J. R. *Nature* **350**, 50–53 (1991).
- Wong, C. S. & Chan, Y.-H. *Tellus* **43B**, 206–223 (1991).

- Brès, B., Brunet, C., Metzl, N., Poisson, A. & Schauer, B. *NATO/ASI Abstr. The Global Carbon Cycle, II Cioccio 8–20 September* (1991).
- Merlivat, L., Etcheto, J. & Boutin, J. *Adv. Space Res.* **11**, 377–385 (1991).
- Enting, I. G. & Mansbridge, J. V. *Tellus* **41B**, 111–126 (1989).
- Keeling, C. D., Piper, S. & Heimann, M. *Am. geophys. Chem. Monogr.* **55**, 305–363 (AGU, Washington DC (1989)).
- Weiss, R. F. *Mar. Chem.* **2**, 203–215 (1974).
- Johnson, K. S. *Limnol. Oceanogr.* **27**, 849–855 (1982).
- Oberhuber, J. *Max Planck Inst. für Met. Rep.* No 15 (1988).
- Wright, P. *Max Planck Inst. für Met. Rep.* **14** (1988).
- Keeling, C. D. & Mansbridge, J. V. *Tellus* **B43**, 156–170 (1991).
- Phillips, L. F. *Geophys. Res. Lett.* **18**, 1221–1224 (1991).

ACKNOWLEDGEMENTS. We acknowledge discussion with C. Goyet, P. S. Liss, L. Merlivat and U. Siegenthaler. This work was carried out as part of the UK BOFS programme, funded by NERC with contributions from National Power.

Phytoplankton productivity in the North Pacific ocean since 1900 and implications for absorption of anthropogenic CO₂

Paul G. Falkowski* & Cara Wilson

Oceanographic and Atmospheric Sciences Division,
Brookhaven National Laboratory, Upton, New York 11973, USA
* Also at: Laboratoire de Physique et Chimie Marines, CNRS-Université
Paris VI, B.P. 08, 06230 Villefranche-sur-mer, France

THE world's carbon budget has not been in steady state since the beginning of the Industrial Revolution¹. At present, carbon dioxide released by anthropogenic activities adds about 7 ± 1.2 gigatonnes (Gt) C yr⁻¹ to the atmosphere, of which about 2 Gt C yr⁻¹ is thought to be sequestered in the oceans². In the steady state, phytoplankton fix about 35–50 Gt C yr⁻¹, representing a significant component of the natural carbon cycle¹. If ocean productivity were changing, these biological processes could have a significant influence on anthropogenic CO₂ levels by drawing down the CO₂ concentration in surface waters and increasing the concentration gradient across the air-sea interface^{1,3,4}. The question of productivity changes is unresolved, however^{2,5,6}. Venrick *et al.*⁷ reported that phytoplankton chlorophyll concentrations had roughly doubled in the central North Pacific gyre between 1965 and 1985. Here we use historical records of Secchi depth data to investigate whether such dramatic changes in phytoplankton biomass have occurred throughout the North Pacific ocean during this century. We find that, although very minor changes may have occurred in this basin over the past 70 years, they are too small to have a significant effect on the rise in atmospheric CO₂ concentrations.

Introduced to oceanography in the nineteenth century by a Jesuit astronomer⁸, the white Secchi disk, between 0.25 and 0.50 m in diameter, is lowered into the ocean and an observer measures the depth (Z_D) at which the disk disappears from view. This simple device is relatively insensitive to changes in observer technique and is still used by oceanographers. The Secchi depth (Z_D) can be related to optical coefficients of sea water through

$$Z_D = \ln[(A - R_{ZD})/R_0 C]/(c + K) \quad (1)$$

where A is the disk albedo, R_{ZD} and R_0 are the reflectance functions at depths Z_D and the surface, C is a function of observer variability, c is the average beam attenuation coefficient (m⁻¹) and K is the average diffuse attenuation coefficient (m⁻¹) for downwelling irradiance.

Gordon and Wouters⁹ and Preisendorfer¹⁰ analysed the errors associated with different terms. Most of these are small, contributing to random, non-systematic errors of less than 10%. The primary source of variability in Secchi disk readings is in the $(c + K)$ term, which covaries with the amount of attenuating material in the water. In open ocean waters, with the exception of small amounts of atmospheric dust, all particles suspended in the water column have a biotic origin. Most of the variability in the absorption of visible light in the open ocean is related to the abundance of phytoplankton, which contain pigments such as chlorophyll¹¹, and Secchi depths can be empirically related to the variability in phytoplankton chlorophyll (Chl)^{12–15} by the equation

$$\text{Chl} = 457 Z_D^{-2.37} \quad (2)$$

(where Chl is in mg m⁻³ and Z_D in m) which is significant at the 99.9% level.

Nolton¹⁴ compared chlorophyll values derived from Secchi depths with those calculated from the ratio of water-leaving radiances at 440 and 550 nm. The latter forms the basis of satellite-derived chlorophyll estimates¹⁶ in open-ocean (case 1¹⁷) waters. The comparison revealed that 89.7% of the variance in surface chlorophyll could be explained using the water-leaving radiance ratios, and 88.2% of the variance could be explained by Secchi depths. Thus, there is statistically only about 1.5% more uncertainty in estimating chlorophyll in the upper open ocean from Secchi depth measurements than from radiance ratio measurements at the sea surface. In fact, global maps of chlorophyll derived from Secchi data are closely comparable to those inferred from satellite measurements of ocean colour¹³.

Secchi depths are the earliest oceanographic measurements of ocean transparency, and the global data set is one of the few historical measurements that can be analysed in biological oceanography from the beginning of the twentieth century. More than 116,000 Secchi depth observations, spanning the years 1900 to 1981, are archived by the National Oceanic Data Center (NODC); most observations are for coastal waters. The data analysed here, for the central ocean, consist of 10,733 observations made since 1907 and were primarily confined to the North Pacific, where the data set is richer both spatially and temporally

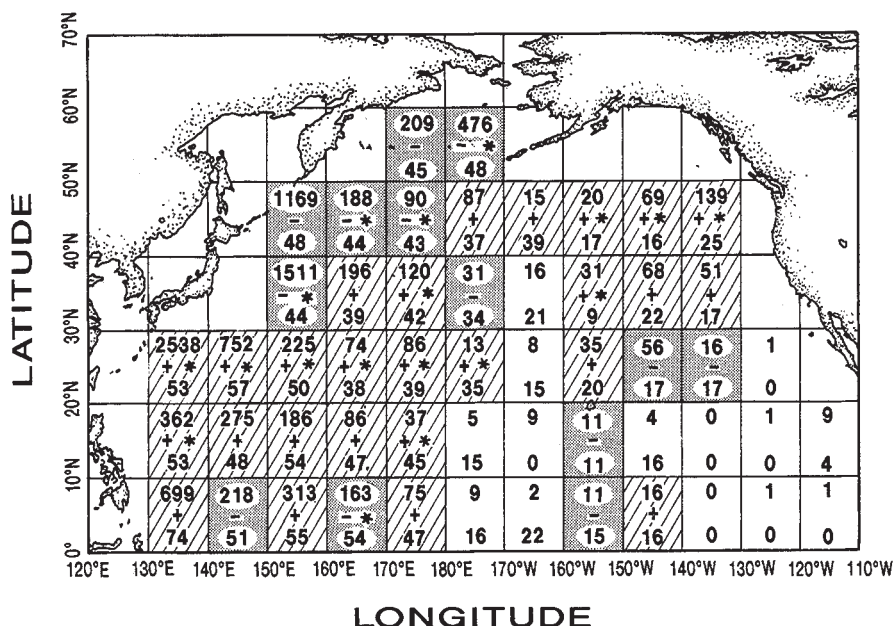


FIG. 1 Spatial distribution of the time-series analysis of Secchi disc data for the North Pacific. Each box corresponds to a 10° × 10° grid. Grids with increasing Secchi depths are dark grey and marked '+', light grey grids correspond to areas with decreasing Secchi depths and are marked '-'. Grids with significant slopes ($P < 0.05$) are marked with an asterisk. The top number in each grid corresponds to the total number of reported observations, and the bottom number corresponds to the years spanned by the data.

TABLE 1 Annual changes in North Pacific Ocean

	Number of 10° × 10° grids	Total number of observations	Average Secchi depth (m)	Average annual change in Secchi depth (m yr ⁻¹)	Average chlorophyll (mg m ⁻³)	Average annual change in chlorophyll (mg m ⁻³ yr ⁻¹)	Average primary production (g C m ⁻² yr ⁻¹)	Average annual change in primary production (g C m ⁻² yr ⁻¹)
All grids	40	10,733	26.1 ± 10.23	0.06 ± 0.29	0.328 ± 0.306	-0.002 ± 0.010	188.7 ± 84.6	-0.59 ± 2.69
Mean and s.d. of all positively sloped grids	26	6,568	30.8 ± 8.30	0.20 ± 0.19	0.25 ± 0.22	-0.006 ± 0.010	166.5 ± 66.2	-1.82 ± 2.31
Mean and s.d. of all negatively sloped grids	13	4,049	18.93 ± 8.53	-0.21 ± 0.27	0.49 ± 0.39	0.006 ± 0.005	233.7 ± 101.3	1.82 ± 1.64

Calculated annual changes in Secchi depths, chlorophyll concentration and primary production for the North Pacific Ocean since the early twentieth century. The grids with increasing and decreasing Secchi depths are shown in Fig. 1. Data from each grid were analysed and the mean and s.d. for each grid were individually calculated.

than other main ocean basins. The data were grouped into 10° square grids and a linear regression of the time-dependent variation in Secchi depths was done individually for each grid with more than 10 observations (Fig. 1). A subset of the grids was tested for assumptions of normality, linearity, homogeneity of variance and independence¹⁸. The results showed that the variance is homogeneously distributed, and that transformations of the data did not improve the normality of the error terms. It should be emphasized that this approach only reveals overall long-term trends; although there could be several increases and decreases within a time span, the number of observations and their distribution in time do not allow for analysis of that resolution. For example, no differences could be discerned between average and El Niño years.

Of the 40 grids with sufficient data for analyses, 26 had an increase in Secchi depth over time, 13 a decrease, and one grid was unchanged (Fig. 1). Thirteen of the 26 linear regressions yielding positive slopes were statistically significant (*F*-test) at the 95% level, whereas five of the 13 negatively sloped grids were significant. The analyses suggest that throughout much of the central North Pacific basin, upper ocean transparency has increased slightly since the beginning of the twentieth century. In grids corresponding to the Kuroshio current and the equatorial Pacific, upper-ocean transparency seems, on average, to have decreased.

Using both the Moran-Geary¹⁸ and Mantel¹⁹ tests, we examined whether the spatial pattern of the changes in the Secchi depths presented in Fig. 1 is random. The results of both tests suggested that although the magnitudes of the slopes are not spatially autocorrelated, the signs of the slopes are, implying that the spatial pattern shown in Fig. 1 is nonrandom. We therefore assume that the differences in the sign of the change in the Secchi depths between different grids are not due to systematic changes in the measurement procedures (such as changes in disk albedo or correction for line angle), but may be due to subtle changes in physical or chemical forcing in the open ocean. In the North Atlantic, changes in Secchi depths are randomly distributed, but the paucity of reported observations from the eastern portion of the North Atlantic Ocean precludes rigorous spatial analysis.

In the North Pacific Ocean, the average Secchi depth is 26.2 ± 10.23 m (mean ± s.d.). The linear regression analyses of the data with time suggest that the average depth is increasing by 0.06 ± 0.29 m per annum (Table 1). This increase in ocean transparency corresponds to an average annual decrease in chlorophyll of 0.002 mg m⁻³ and is not significantly different from zero (*P* < 0.05). Analysing only those grids with significant positive slopes, we find that the average change in the Secchi depth is 0.20 ± 0.19 m, which corresponds to a decrease in chlorophyll of 0.006 ± 0.010 mg m⁻³ yr⁻¹. Such small changes are at the limit of more modern chlorophyll analysis methods, which have been in common use only since the mid-1960s.

Although the calculated changes in chlorophyll are only representative of the upper portion of the water column and are less than 2% of the mean, the surface chlorophyll concentrations can be used statistically to derive integrated values and primary production²⁰.

We calculated primary production rates by converting the Secchi depths values to chlorophyll through equation (2), and converting the chlorophyll to primary production (PP) rates using the relationship described by Eppeley *et al.*²¹

$$PP \text{ (mg C m}^{-2} \text{ d}^{-1}\text{)} = 1,000 [\text{Chl (mg m}^{-3}\text{)}]^{1/2} \quad (3)$$

where Chl is the concentration of chlorophyll in the near surface ocean. As equations (2) and (3) are both nonlinear, chlorophyll and primary production values were calculated for the initial and final date recorded in each grid. The difference, divided by the years spanned by the data, yields an average annual value. The calculated annual mean primary production in the North Pacific is 188.7 g C m⁻² and the average annual rate of change is -0.59 g C m⁻². Regions with increasing Secchi depths have an average productivity of 166.5 g C m⁻² yr⁻¹, and regions with decreasing Secchi depths average 233.7 g C m⁻² yr⁻¹ (Table 1). The mean values of primary production calculated from Secchi depths are surprisingly comparable to recent radiocarbon measurements in the North Pacific^{22,23}, but because of inherent variance, the calculated annual rates of change are much too small to be reliably measured with any known oceanographic technique, even over decadal timescales. Our calculations suggest that the average total primary production for the North Pacific, representing an area of 4.94 × 10¹³ m², is 9.3 Gt C yr⁻¹, and is decreasing at an average annual rate of 0.03 Gt C, which is not significantly different from zero.

It might not be surprising that phytoplankton biomass and primary production in a principal central ocean basin have not changed significantly in this century; CO₂ is not limiting carbon fixation in marine systems, and there is no obvious fertilization of the open oceans. Venrick *et al.*⁷ suggested that the apparent doubling of chlorophyll they reported for the central North Pacific gyre might be due to increased storm activity or intensity, which effectively increased the flux of essential plant nutrients from below the pycnocline to the euphotic zone. Our analysis indicates that although phytoplankton biomass and productivity may have changed in parts of the North Pacific Ocean, the changes are much smaller and of opposite sign in the region studied by Venrick *et al.*

Does phytoplankton carbon fixation contribute significantly to the flux of excess atmospheric CO₂ into the ocean? To account for all of the estimated ocean uptake of 2 Gt C yr⁻¹, net phytoplankton productivity would have to increase exponentially by about 4% each year (assuming that total global production is about 50 Gt C yr⁻¹), and the fixed carbon would have to be sequestered either below the main pycnocline or in the sediments (export production²³). An increase in production would require

either the addition of nutrients external to the ocean²⁴, or an increase in the efficiency of their use in areas where nutrients are in excess. Throughout most of the central ocean basins, inorganic fixed nitrogen and phosphate are vanishingly small in the euphotic zone. In these areas, increased turbulent mixing within the ocean could lead to increased upwelling of nutrients from below the pycnocline, but such an upwelling flux would include inorganic carbon. The potential stimulation of primary production induced by such forcing would have no significant effect on the atmospheric concentration of CO₂; the ratio of inorganic carbon to other essential inorganic nutrients (the Redfield ratio) below the pycnocline is virtually identical to the elemental composition of phytoplankton².

There are three main open-ocean regions where inorganic nutrients are in excess in the euphotic zone: the equatorial and subarctic Pacific, and the Southern Ocean^{25,26}. In these three regions it has been argued that phytoplankton productivity is limited by iron, which is supplied to the upper ocean primarily from atmospheric deposition^{25,26}. In principle, the spread of deserts and/or increase of wind speeds could promote the transport of iron-rich particles seawards, and on geological timescales this seems to have stimulated primary production enough to affect atmospheric CO₂ (ref. 26). The spatial pattern of minute changes in Secchi depths in the North Pacific suggests there has been an increase in phytoplankton biomass in the subarctic and equatorial Pacific in this century. This spatial pattern of temporal changes (and lack thereof in the North Atlantic) is consistent with a hypothetical stimulation of phytoplankton biomass resulting from an increased deposition of aeolian iron. Our analysis shows, however, that even if this enhancement occurred, it has had relatively little impact on phytoplankton productivity. Moreover, model calculations suggest that even if all of the excess nutrients in these two areas were consumed by phytoplankton, the net invasion of atmospheric CO₂ would be small²⁷. Because proxy historical data bases, such as Secchi depths, are extremely sparse or non-existent for other main ocean basins (such as the Southern Ocean), we cannot preclude the possibility that changes in primary production have occurred elsewhere. But if our analysis is representative of other ocean basins, it seems reasonable to assume that open-ocean phytoplankton biomass and production have not changed significantly since the beginning of the Industrial Revolution. Consequently, for the purposes of modelling the sources and sinks of excess atmospheric CO₂, open-ocean biological processes can be considered to be in a steady state^{2,5,27}.

In contrast to the open ocean, coastal ocean regions have experienced large increases in nutrient supply correlated with increases in human population levels in coastal areas²⁸, and it has been suggested that this enrichment is large enough significantly to affect anthropogenic CO₂ levels²⁹. If that were the case, we would expect to see changes in phytoplankton biomass on continental shelves and a carbon sink amounting to ~2 Gt C yr⁻¹ in either the interior ocean or the sediments of the continental margins. Unfortunately, although there are abundant Secchi depth data for continental margins, the data are difficult to interpret using equation (2) because of the interference of terrigenous particles and other absorbing materials. Nonetheless, there is compelling evidence that phytoplankton biomass and, presumably, productivity, have been increasing on continental shelves throughout the twentieth century³⁰. As yet there is no conclusive evidence for a significant sink of recently produced carbon in the coastal oceans³¹; nearly all organic matter seems to be efficiently reoxidized in the upper 500 m of the ocean^{32,33}. □

3. Watson, A. J., Robinson, C., Robinson, J. E., Williams, P. LeB. & Fasham, M. J. R. *Nature* **350**, 50–53 (1991).
4. Smith, W. O. Jr et al. *Nature* **352**, 514–516 (1991).
5. Peng, T.-H. & Broecker, W. S. *J. geophys. Res.* **89**, 8170–8180 (1984).
6. Smith, S. V. & MacKenzie, F. T. *Global biogeochem. Cycles* **5**, 189–190 (1991).
7. Venrick, E. L., McGowan, J. A., Cayan, D. R. & Hayward, T. L. *Science* **238**, 70–72 (1987).
8. Secchi, A. *Sul Moto Ondoso del Mare e su le Correnti di esso Specialmente Auquale Littorali*, 2nd Edn, 258–288 (1886). (Engl. transl.) Department of the Navy, Office of Chief of Naval Operations, ONI Transl. A-655, Op-923 M4B.
9. Gordon, H. R. & Wouters, A. W. *Appl. Opt.* **17**, 3341–3343 (1978).
10. Preisendorfer, R. W. *Limnol. Oceanogr.* **31**, 909–926 (1986).
11. Kirk, J. T. O. *Light and Photosynthesis in Aquatic Ecosystems* (Cambridge Univ. Press, New York, 1983).
12. Simonot, J. Y. & Le Treut, H. *J. geophys. Res.* **91**, 6642–6646 (1986).
13. Lewis, M. R., Kuring, N. & Yentsch, C. *J. geophys. Res.* **93**, 6847–6856 (1988).
14. Nolten, J. W. thesis, California State College at Moss Landing (1980).
15. Megard, R. O. & Berman, T. *Limnol. Oceanogr.* **34**, 1640–1655 (1989).
16. Gordon, H. R., Clark, D. K., Mueller, J. L. & Hovis, W. A. *Science* **210**, 63–66 (1980).
17. Morel, A. & Prieur, L. *Limnol. Oceanogr.* **22**, 709–722 (1977).
18. Cliff, A. D. & Ord, J. K. *Spatial Processes: Models and Applications* (Pion, London, 1981).
19. Mantle, N. *Cancer Res.* **27**, 209–220 (1967).
20. Morel, A. & Berthon, J.-F. *Limnol. Oceanogr.* **34**, 1545–1562 (1989).
21. Eppley, R. W., Stewart, E., Abbott, M. R. & Heyman, U. J. *Plankt. Res.* **7**, 57–70 (1985).
22. Chavez, F. P. & Barber, R. T. *Deep-Sea Res.* **34**, 1229–1243 (1987).
23. Berger, W. H., Fischer, K., Lai, C. & Wu, G. *Ocean Productivity and Organic Carbon Flux, Part I* *Scripps Inst. Oceanogr. Ref. 87-30* (Univ. California, San Diego, La Jolla, 1987).
24. Barber, R. T. in *Primary Productivity and Biogeochemical Cycles in the Sea* (eds Falkowski, P. G. & Woodhead, A. D.) 89–106 (Plenum, New York, 1992).
25. Martin, J. H. in *Primary Productivity and Biogeochemical Cycles in the Sea* (eds Falkowski, P. G. & Woodhead, A. D.) 123–136 (Plenum, New York, 1992).
26. Martin, J. H. *Paleoceanography* **5**, 1–13 (1990).
27. Sarmiento, J. & Orr, J. *Limnol. Oceanogr.* **36**, 1928–1950 (1991).
28. Peierls, B. L., Caraco, N. F., Pace, M. L. & Cole, J. J. *Nature* **350**, 386–387 (1991).
29. Walsh, J. J. *Nature* **350**, 53–56 (1991).
30. Walsh, J. J. *On the Nature of Continental Shelves* 520 (Academic, New York, 1988).
31. Anderson, R. F. et al. *Cont. Shelf Res.* (in the press).
32. Rowe, G. T. et al. *Nature* **324**, 559–561 (1986).
33. Falkowski, P. G. et al. *Cont. Shelf Res.* **8**, 457–484 (1988).

ACKNOWLEDGEMENTS. We thank D. Wallace, J. Sarmiento, W. Broecker, T. Platt, W. Broenkow, A. Bakun and A. Morel for comments, and N. Oden and W. Medeiros for advice on statistical analysis. This research was supported by the Electric Power Research Institute, the Centre National de la Recherche Scientifique, and the John Simon Guggenheim Memorial Foundation.

Experimentally produced oscillatory zoning in the (Ba, Sr)SO₄ solid solution

Andrew Putnis*, Lurdes Fernandez-Diaz† & Manuel Prieto‡

* Department of Earth Sciences, University of Cambridge, Cambridge CB2 3EQ, UK

† Departamento de Cristalografía y Mineralogía, Universidad Complutense de Madrid, 28040 Madrid, Spain

‡ Departamento de Geología, Universidad de Oviedo, 33005 Oviedo, Spain

WHEN crystals grow from a multicomponent fluid phase under conditions where ionic diffusion in the solid is negligible compared to that in the liquid, any compositional gradients in the crystal record the evolution of the solid/liquid interface composition during growth. For the particular case of oscillatory zoning^{1–3}, a relatively common feature of natural crystal growth⁴, there has been considerable theoretical interest^{4–6}, but the specific question of whether high or low supersaturations are required to explain the development of the zoning remains unanswered. Experimentally produced compositional oscillations have been observed^{7,8}, but the role of supersaturation had to be inferred, rather than measured directly. Here we describe major-element oscillatory zoning in (Ba, Sr)SO₄ solid solutions grown by the counter-diffusion of (Ba²⁺, Sr²⁺) and SO₄²⁻ ions through a porous silica-gel transport medium. We demonstrate how the different solubilities of the two pure phases determine the threshold supersaturation for nucleation, and show how coupling between the compositional gradients in the solid and the liquid results in the observed oscillatory behaviour.

In the experiment (Fig. 1), the silica gel acts as a transport medium, allowing counter-diffusion of the anionic and cationic reactants, and effectively suppressing convective transport. Diffusion-controlled concentration profiles of Ba, Sr and S and

Received 28 October 1991; accepted 14 July 1992.

1. Intergovernmental Panel on Climate Change Working Group I. *Climate Change, the IPCC Scientific Assessment* (Cambridge Univ. Press 1990).
2. Sarmiento, J. L. in *Primary Productivity and Biogeochemical Cycles in the Sea* (eds Falkowski, P. G. & Woodhead, A. D.) 317–332 (Plenum, New York, 1992).

of pH develop in the gel column. By measuring the concentration of these components through the gel column as a function of time, and using a Debye-Hückel model of ionic complexing in the aqueous solution, the activities of the components and the evolution of the supersaturation through the silica-gel column have been calculated⁹. Under these conditions nucleation takes place at high values of the supersaturation, the actual value of the threshold supersaturation required for nucleation depending on the initial boundary conditions (the initial concentration of the reactants in tubes and the length of the gel column). These factors determine the concentration gradients, which in turn control the supersaturation rate (the rate at which the system is moving out of equilibrium). Because of the large difference in the solubilities of BaSO_4 and SrSO_4 , the threshold values of the supersaturation of $(\text{Ba}, \text{Sr})\text{SO}_4$ are strongly dependent on the composition. Crystals of $(\text{Ba}, \text{Sr})\text{SO}_4$ nucleate within the silica gel at a point in the tube where the threshold supersaturation is exceeded and the supersaturation rate is a maximum, and then continue to grow for times up to several months. At the end of the experiment, the millimetre-sized crystals are removed from the gel and sectioned for scanning electron microscopy and microanalysis.

A backscattered scanning electron micrograph of a section through a typical crystal (Fig. 2a) shows a coarse compositional zoning from core to rim, between strontium-rich (core) and barium-rich regions ($\text{Sr}_{0.75}\text{Ba}_{0.25}\text{SO}_4$ to $\text{Sr}_{0.10}\text{Ba}_{0.90}\text{SO}_4$). This cycle is repeated twice, but superimposed on it are many oscillations with smaller wavelength and compositional amplitude. In some orientations the crystals show the development of growth sectors (Fig. 2b) in which the different contrast indicates a crystallographic control of the Ba/Sr ratio of the sulphate precipitating on each face (sector zoning). The overall cyclic sequence from strontium-rich to barium-rich is similar in the two crystals, indicating a 'crystal stratigraphy' which records the changing physical-chemical environment during growth.

The equilibrium between a solid solution and an aqueous solution is described by the Lippmann model¹⁰ in which the coexisting compositions are represented by solidus and solutus lines respectively on a phase diagram. Such a diagram has been calculated for the system $(\text{Ba}, \text{Sr})\text{SO}_4\text{-H}_2\text{O}$ (Fig. 3). The much greater solubility of SrSO_4 leads to the result that even from a Sr-rich aqueous solution, the equilibrium precipitate is almost pure BaSO_4 . Only when the aqueous solution is virtually barium-free will a strontium-rich sulphate precipitate. These general conclusions are only slightly affected by deviations from ideality in the $(\text{Ba}, \text{Sr})\text{SO}_4$ solid solution (M.P., A.P. and L.F.-D., manuscript in preparation).

The diffusion rates of Ba^{2+} and Sr^{2+} through the silica gel are similar¹² and at the point where the first crystals nucleate, the Lippmann diagram (Fig. 3) predicts that the first solid phase to precipitate should be virtually pure BaSO_4 . The experimental observations show, however, that the cores of the crystals are

strontium-rich. This can be explained by the effect of supersaturation.

Nucleation and growth in this experimental system takes place under non-equilibrium conditions in which the concentrations in the aqueous solution are time- and space-dependent⁹. The threshold value of the supersaturation required to nucleate a phase depends not only on its solubility, which defines the concentration of ions in solution for any given value of the supersaturation, but also on the rate at which the system is moving out of equilibrium⁹ (the supersaturation rate). Thus very high supersaturations can be achieved without any nucleation in solutions of sparingly soluble salts. Experiments have shown (M.P., A.P. and L.F.-D., manuscript in preparation) that the threshold values of supersaturation for the experiments described here are around 10,000 for BaSO_4 compared with 100 for SrSO_4 .

The result is that although the aqueous solution may be much more supersaturated with respect to BaSO_4 than SrSO_4 , the nucleation of a strontium-rich phase may be kinetically preferred because of its lower supersaturation threshold, β^* . In an aqueous solution in which the concentration of reactants is increasing with time, the composition of the solid solution that nucleates is that for which the threshold supersaturation is first exceeded. Changing the start boundary conditions (the initial compositions in the reservoir solutions or the length of the gel column) changes the supersaturation rate and hence the composition of the first nuclei.

The non-equilibrium nucleation of a strontium-rich phase depletes the surrounding aqueous solution in Sr^{2+} and sets up compositional gradients. The time-dependence of these gradients is controlled by the diffusion rate in the solution, the partitioning of Ba^{2+} and Sr^{2+} between crystal and solution, and the growth rate of the crystal. Although we cannot at present specify the exact nature of the coupling between the composi-

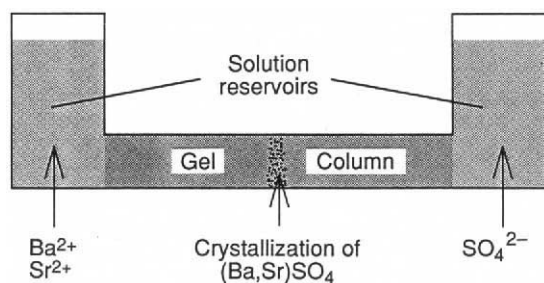
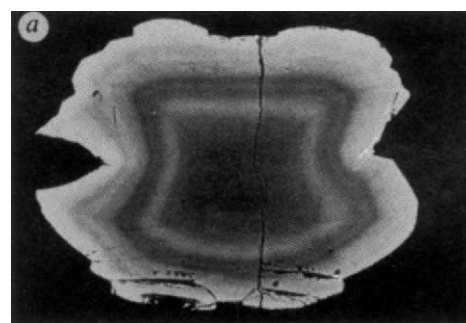
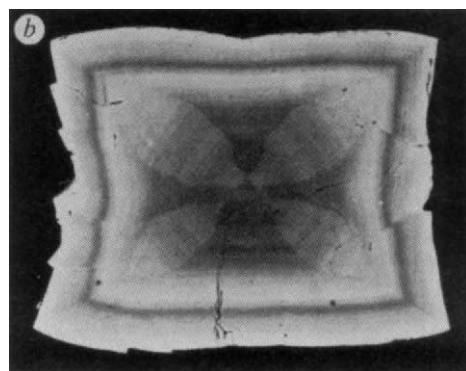


FIG. 1 Schematic representation of the experimental U-tube arrangement. The role of the silica gel is to allow diffusive transport while suppressing convective transport of the reacting components, thus guaranteeing diffusional control of the crystal growth process. The concentrations of the parent solutions used were: Ba^{2+} , 0.3 M; Sr^{2+} , 0.3 M; and SO_4^{2-} , 0.5 M.



20 kV 10 μm WD24



20 kV 10 μm WD24

FIG. 2 a, Backscattered scanning electron micrograph of a section through a typical crystal. Dark regions are strontium-rich, whereas white regions are barium-rich. b, Similar electron micrograph, but in an orientation showing growth sectors.

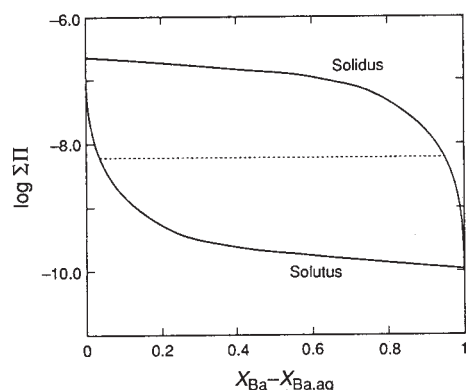


FIG. 3 Calculated Lippmann diagram for the BaSO_4 - SrSO_4 solid solution. Unlike traditional phase diagrams, $\Sigma\Pi$ values for the solid and the liquid are plotted on the ordinate against two superimposed scales on the abscissa: X_{Ba} refers to the concentration of Ba in the solid, and $X_{\text{Ba,aq}}$ refers to the concentration of Ba in the aqueous solution. The dashed tie line shows the equilibrium compositions of the coexisting solid (on the solidus line) and the aqueous solution (on the solutus line). $\Sigma\Pi$ is the total solubility product defined as $[\text{SO}_4^{2-}][\text{Ba}^{2+}] + [\text{Sr}^{2+}]$ where the quantities in the square brackets are the activities of these ions in the aqueous phase. The equilibrium values of $\Sigma\Pi$ for the coexisting solid and the liquid phase are related by the thermodynamic solubility products of the pure endmembers¹⁰.

tional gradients in the solid and the liquid phase, there are several possible feedback mechanisms. It has been suggested^{7,12} that the partition coefficients between a solid and liquid at the interface depend on the growth rate. But the growth rate is itself dependent on the supersaturation which is defined by the value of these partition coefficients. The threshold supersaturation required for growth is also likely to depend on the composition of the solid. Finally, considerable strain energy is associated with changing the composition of the growing crystal, because of the 12.6% increase in volume between the SrSO_4 and BaSO_4 .

The overall result is a type of autocatalytic surface attachment^{4,6}, in which the growth of a strontium-rich composition increases the relative concentration of barium in the aqueous solution until the threshold supersaturation for the growth of a more barium-rich composition is exceeded. This in turn depletes the liquid in barium, resulting in a period of more strontium-rich growth. The overall compositional change from the core of the crystal is towards barium enrichment, driven by the high values of the supersaturation for the BaSO_4 endmember, but is subject to small-scale oscillatory adjustments to the composition of the growing interface. These reflect the competition between the thermodynamics (favouring barium enrichment) and the kinetics (favouring strontium enrichment because of its lower value of β^*). When the composition reaches around $\text{Ba}_{0.9}\text{Sr}_{0.1}\text{SO}_4$ (the growth zone with the most barium-rich composition), the barium depletion in the liquid drives the system back to the beginning again, with the growth of a strontium-rich composition and the start of another main cycle.

The analysis of the coupling between the compositional gradients in the crystal and those in the aqueous solution may be simplified because it is not necessary to consider explicitly all of the factors that define this dependence. The composition profiles of barium and strontium in the crystal record the evolution of the interface composition during growth, and because these compositional gradients in the solid are coupled to each other by the liquid phase, the system can be fully described without explicitly considering the aqueous solution. The probability of incorporating a Ba^{2+} or Sr^{2+} ion into the growing surface of the solid solution depends on the concentration of that component in the liquid phase (the supersaturation) and on the spatial gradient of this concentration. In its simplest form the coupling can be described¹³ by a generalized Ginzburg-Landau

rate equation of the type

$$\frac{dQ_1}{dt} = -x_1 Q_1 + \lambda_1 \frac{dQ_2}{dz} \quad \frac{dQ_2}{dt} = -x_2 Q_2 + \lambda_2 \frac{dQ_1}{dz}$$

where Q_1 and Q_2 are order parameters which can be equated to the composition in the crystal, of Ba and Sr, respectively, t is time and z is the space coordinate. λ_1 and λ_2 are constants related to the growth rate, and x_1 and x_2 are inverse time constants. For specific boundary conditions, the solution to such an equation results in an oscillatory variation of Q_1 and Q_2 in space and time, described by a first-order Bessel function¹³.

Experiments of the type described here can test the factors controlling oscillatory crystal growth. We have described elsewhere¹⁴ the similarity in the morphological development of BaSO_4 crystal growth in gels and in nature, and suggested that this technique may therefore be able to reproduce natural growth processes. Furthermore, by suppressing convective transport, we also simulate crystal growth from a free solution under microgravity conditions^{8,15}. The experiments that we report here have shown that oscillatory zoning can take place in a solid solution crystallizing under supersaturated conditions if the threshold supersaturation for nucleation and growth is strongly dependent on composition, a situation which will exist if the endmembers have significantly different solubility. Oscillatory zoning does not necessarily require a non-ideal solid solution with a miscibility gap, as suggested by some models, nor an externally controlled alternation in intensive parameters^{2,3}. □

Received 24 February; accepted 14 July 1992.

1. Allégre, C. J., Provost, A. & Jaupart, C. *Nature* **294**, 223-228 (1981).
2. Jamveit, B. *Am. Miner.* **76**, 1319-1327 (1991).
3. Yardley, B. W. D., Rochelle, C. A., Barnicoat, A. C. & Lloyd, G. E. *Min. Mag.* **380**, 357-366 (1991).
4. Ortoleva, P. *Earth Sci. Rev.* **29**, 3-8 (1990).
5. Nicolis, G. & Prigogine, I. *Self-Organization in Non-Equilibrium Systems* (Wiley, New York, 1977).
6. Ortoleva, P., Merino, E., Moore, C. & Chadam, J. *Am. J. Sci.* **287**, 979-1007 (1987).
7. Reeder, R. J., Fagioli, R. O. & Meyers, W. J. *Earth Sci. Rev.* **29**, 39-46 (1990).
8. Hurlé, D. T. J. *Adv. Space Res.* **3**, 43-50 (1983).
9. Prieto, M., Putnis, A. & Fernandez-Diaz, L. *Geol. Mag.* **127**, 485-495 (1990).
10. Glynn, P. D. & Reardon, E. J. *Am. J. Sci.* **290**, 164-201 (1990).
11. Chernov, A. A. *Modern Crystallography III: Crystal Growth* (Springer, Berlin, 1984).
12. Prieto, M., Fernandez-Diaz, L. & López-Andrés, S. *J. Cryst. Growth* **98**, 447-460 (1989).
13. Salje, E. & Wruck, B. *Phys. Chem. Miner.* **16**, 140-147 (1988).
14. Prieto, M., Putnis, A., Arribas, J. & Fernandez-Diaz, L. *Miner. Mag.* (in the press).
15. Goldsmith, P. J. *J. Cryst. Growth* **79**, 37-42 (1986).

ACKNOWLEDGEMENTS. We thank E. Salje and B. Wruck for discussions. This work was partially supported by CICYT (Science and Technology Commission of Spain) and Mobil North Sea Limited.

Evidence from ultra-high-pressure marbles for recycling of sediments into the mantle

Harry Becker* & Rainer Altherr†

* Max Planck Institut für Chemie, Abteilung Geochemie, Saarstrasse 23, D-W-6500 Mainz, Germany

† Institut für Petrographie und Geochemie, Universität Karlsruhe, Kaiserstrasse 12, D-W-7500 Karlsruhe 1, Germany

ROCKS of crustal origin metamorphosed at ultra-high pressures ($P > 2.5$ GPa) have been described from several orogenic belts¹⁻⁵. In the western Alps, for example, ultra-high pressure rocks originally equilibrated at about 750 °C and 3.5 GPa (ref. 1). During decompression, these rocks were cooled considerably and therefore did not pass through the granulite stability field. Here we describe ultra-high-pressure metasediments that have equilibrated at higher temperatures ($> 1,100$ °C) and followed a different exhumation path. A calcsilicate marble from the Bohemian massif contains clinopyroxenes with potassium-rich feldspar exsolutions. Potassium contents in the original clinopyroxenes indicate crystallization at pressures above 3-4 GPa (refs 6, 7). The high peak

pressures and temperatures inferred for this rock, and its association with high-temperature peridotites and high-pressure granulites, suggest that carbonate sediments were subducted into the upper mantle, equilibrated at mantle conditions and were then emplaced in the crust along with mantle rocks. Our observations thus support suggestions based on less direct evidence (such as mass-balance considerations in orogenic belts⁸, and ocean-island basalt geochemistry^{9,10}) that a limited amount of sediment must be recycled into the mantle.

The Moldanubian association of high-pressure granulites with high-temperature peridotites in Lower Austria (southern Bohemian Massif) forms part of the internal crystalline zone of the Palaeozoic Variscan collision belt. Some aspects of the pressure-temperature (P - T) history of the granulites and peridotites have been worked out recently¹¹⁻¹⁹. Calcsilicate marbles found in this rock association contain clinopyroxenes with exsolution lamellae of potassium-rich alkali feldspar and therefore provide evidence for an equilibration at extreme P - T conditions. The calcsilicate marbles occur as blocks a few tens of metres distant from a serpentinite outcrop in the Dunkelsteiner Wald area (~60 km west of Vienna). Garnet pyroxenite 'nodules' are enclosed in a matrix of calcite, porphyroclastic clinopyroxene (cpx), cpx neoblasts, rare garnet with inclusions of sphene and cpx (occasionally accompanied by alkali feldspar), and secondary hornblende, alkali feldspar, plagioclase and biotite. Porphyroclastic cpx in the nodules and in the matrix contains abundant exsolution lamellae or blebs of alkali feldspar together with small grains of sphene (Figs 1, 2a). Cpx neoblasts in the nodules lack exsolutions and coexist with garnet, sphene and alkali feldspar in recrystallized domains or are found around cpx porphyroclasts. A third cpx generation is intergrown with secondary plagioclase and amphibole, forming symplectites which occasionally surround garnet. Sphene grains enclosed in garnet and cpx contain lamellae of a high- T amphibole (Fig. 2b), an unusual feature which is not known from other sphene occurrences.

Porphyroclastic cpx grains in nodules and matrix show strong chemical zoning (Table 1). A pronounced increase in aluminium from core to rim is accompanied by a decrease in sodium. Zoning in cpx neoblasts is different. They show high-Al, low-Na cores, with constant or slightly decreasing Al and decreasing Na towards the rims. Cores of porphyroclastic cpx grains are low in titanium whereas their rims and the neoblasts have elevated

abundances. Some cores of porphyroclastic cpx have low potassium contents (up to 0.07% K_2O in matrix and nodules, which is close to the detection limit of routine electron microprobe analysis). Most significantly, some cpx inclusions in sphene have retained K_2O contents of up to 0.33 wt% (Table 1b). Garnets in nodules and matrix are zoned with calcium decreasing and iron increasing from core to rim (Table 1a and b).

Unequivocal constraints for an exsolution origin of the alkali feldspar lamellae in cpx porphyroclasts are provided by electron microprobe traverses. Figure 3 shows an example of such a profile across the cpx porphyroclast shown in Fig. 1. The cpx host shows strong chemical zoning near the alkali feldspar lamellae. A decrease in sodium and aluminium (and to a lesser extent in silicon) in cpx near the lamellae is accompanied by an increase in calcium and magnesium. This kind of zoning indicates a diffusive loss of jadeite component to form feldspar with a simultaneous increase of the diopside component in cpx. Although, as stated above, almost no potassium remained in cpx, it is conceivable that the K-feldspar component in the alkali feldspar lamellae formed by exsolution too. Exsolution of the silica-rich phase feldspar has to be balanced by the coupled exsolution of a low-silica phase such as sphene (Fig. 1). Similar balanced exsolution reactions are also reported from other pyroxene occurrences²⁰. Aluminium zoning in the cpx core close to the alkali feldspar lamellae is the opposite of the aluminium zoning that can be observed at the cpx rims close to garnet (Fig. 3). These observations clearly indicate that changes in the P - T conditions cause the observed mineral zoning and the formation of these lamellae. A metasomatic origin of this intergrowth would clearly be incompatible with the observed chemical zoning and also with the fact that the lamellae are restricted to the cores of the earliest cpx generation (Fig. 1). The lamellae are never found in younger cpx neoblasts or symplectites.

Temperature estimates (at an assumed pressure of 2 GPa; pressure dependence of thermometer is $\sim 35^\circ C GPa^{-1}$), based on the partitioning of iron and magnesium between garnet and cpx²¹, give 1,070–850 °C for contacts between host garnet and inclusions of low-potassium cpx and 1,100–970 °C for grain cores. These temperatures are interpreted as cooling temperatures after decompression. Although it is not possible to estimate temperatures for the former ultra-high-pressure stage documented by the K-feldspar exsolution lamellae in the cores of primary cpx, it is conceivable that temperatures were similar or even higher than for the 'low-pressure' re-equilibration.

Minimum pressures for the ultra-high-pressure stage may be deduced from an estimate of the minimal original K_2O content in the primary cpx. The ubiquitous exsolutions of potassium-rich

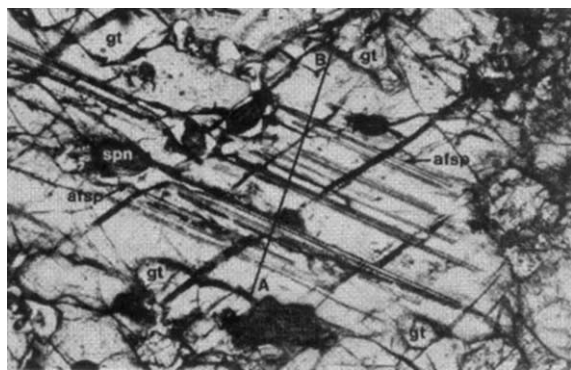


FIG. 1 Cpx porphyroclast from a garnet pyroxenite nodule with exsolution lamellae of alkali feldspar (afsp) which seem to be parallel to one of the cpx cleavage planes. In a few sections (not shown) the lamellae were cut perpendicular to their long dimension. In such cases their rectangular shapes indicate that the lamellae are not planar but lath-shaped. The lamellae do not continue to the porphyroclast rim. Near inclusions of sphene (spn), the afsp exsolutions grade into irregular patches. Garnet (gt) seems to have grown at the expense of cpx. The line A-B represents electron microprobe traverse shown in Fig. 3. Width of micrograph is 3.3 mm.

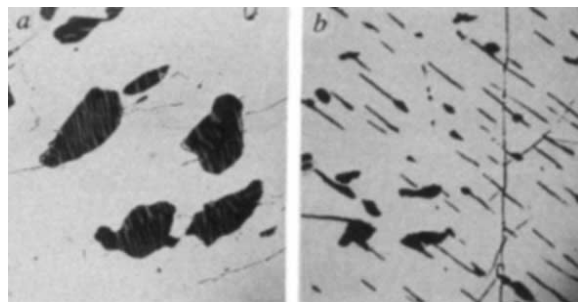


FIG. 2 a, Back-scattered electron image showing exsolution blebs of perthitic alkali feldspar in core of porphyroclastic clinopyroxene. The blebs may form directly by exsolution or by necking and recrystallization of straight alkali feldspar exsolution lamellae. Width of micrograph is 170 μm . b, Back-scattered electron image showing exsolution of amphibole (dark) in sphene enclosed in garnet. Width of micrograph is 80 μm .

TABLE 1 Representative mineral compositions

a, Marble matrix														
	Garnet		Cpx-pcl		Cpx-nbl		Sphe ^{ne*}		Sphe ^{ne†}	Amphibole	Biotite	afsp		Calcite [‡] A33
	Core	Rim	Core	Rim	Core	Rim	Core	Rim						
SiO ₂	39.54	39.22	53.91	47.07	44.40	45.50	30.44	29.89	29.91	37.75	36.07	65.81	63.76	—
TiO ₂	0.44	0.38	0.20	1.01	1.35	1.73	36.06	38.51	38.96	4.40	4.21	—	—	—
Al ₂ O ₃	21.54	21.61	3.84	10.47	14.66	11.72	3.36	1.89	1.46	16.94	15.27	19.30	18.88	—
Cr ₂ O ₃	0.02	—	—	—	—	—	—	—	0.03	0.06	—	—	—	—
Fe ₂ O ₃	—	—	—	—	—	—	0.34	0.47	0.53	—	1.76	—	0.14	—
FeO	11.74	13.67	4.51	7.52	7.27	7.60	—	—	—	13.67	14.26	—	—	0.75
MnO	0.35	0.50	0.03	—	—	0.23	0.03	0.03	0.02	0.08	0.03	—	—	—
MgO	4.12	4.41	12.80	9.55	8.15	8.91	—	—	—	8.33	13.30	—	—	0.76
CaO	22.12	20.11	22.20	22.59	22.70	22.84	29.11	28.83	28.87	12.03	0.10	0.33	0.26	56.75
Na ₂ O	0.04	0.03	1.79	0.91	0.96	0.86	0.03	0.02	0.03	1.79	0.22	2.59	1.43	—
K ₂ O	—	—	0.04	—	—	—	—	—	—	2.18	9.68	12.87	14.67	—
Total	99.91	99.93	99.32	99.33	99.49	99.39	99.35	99.64	99.81	97.23	94.91	100.90	99.14	58.26
Oxygens	12	12	6	6	6	6	5	5	5	23	11	8	8	3
Si	3.006	2.993	1.980	1.766	1.663	1.712	0.997	0.980	0.980	5.775	2.723	2.971	2.964	—
Ti	0.025	0.022	0.006	0.028	0.029	0.049	0.888	0.949	0.960	0.506	0.239	—	—	—
Al	1.930	1.944	0.166	0.463	0.647	0.520	0.130	0.073	0.057	3.054	1.358	1.026	1.035	—
Cr	0.001	—	—	—	—	—	—	—	0.001	0.007	—	—	—	—
Fe ³⁺	—	—	—	—	—	—	0.008	0.012	0.013	—	0.100	—	—	—
Fe ²⁺	0.746	0.873	0.139	0.236	0.228	0.239	—	—	—	1.749	0.901	—	0.005	0.010
Mn	0.022	0.032	0.001	—	—	0.007	0.001	0.001	0.001	0.011	0.002	—	—	—
Mg	0.467	0.502	0.701	0.534	0.455	0.500	—	—	—	1.899	1.496	—	—	0.018
Ca	1.802	1.645	0.874	0.908	0.911	0.921	1.021	1.012	1.013	1.972	0.008	0.016	0.013	0.971
Na	0.006	0.005	0.128	0.066	0.069	0.063	0.002	0.001	0.002	0.559	0.032	0.226	0.129	—
K	—	—	0.002	—	—	—	—	—	—	0.426	0.932	0.741	0.870	—
Cations	8.005	8.016	3.996	4.008	4.011	4.011	3.047	3.028	3.026	15.957	7.791	4.980	5.016	0.999

b, Pyroxenite nodule														
	Garnet [§]		Cpx-pcl		Cpx-nbl		Cpx in Sphe ^{ne}		Sphe ^{ne}		Amphibole	Alkali feldspar		
	Core	Rim	Core	Rim	Core	Rim	Sphe ^{ne}	Core	Rim					
SiO ₂	38.73	38.02	53.36	45.32	40.44	40.27	40.64	30.31	30.05	37.47	65.72	65.52	—	—
TiO ₂	0.38	0.27	0.20	0.91	3.76	3.78	8.10	37.48	38.75	4.20	—	—	—	—
Al ₂ O ₃	21.46	21.36	5.86	13.65	16.66	15.67	12.53	2.60	1.70	16.65	19.03	18.62	—	—
Cr ₂ O ₃	—	—	—	—	—	—	0.03	0.05	0.04	—	—	—	—	—
Fe ₂ O ₃	1.78	1.84	—	—	1.71	3.29	—	0.32	0.48	—	0.12	0.25	—	—
FeO	12.71	15.94	5.31	6.70	5.41	4.78	7.20	—	—	14.12	—	—	—	—
MnO	0.40	1.44	0.08	—	—	0.20	0.13	0.11	—	0.26	—	—	—	—
MgO	5.63	4.93	11.59	8.73	7.28	7.40	7.99	—	—	8.45	—	—	—	—
CaO	18.13	15.19	20.97	22.88	23.28	22.69	21.99	29.33	29.11	11.66	0.67	0.17	—	—
Na ₂ O	0.06	0.01	2.35	0.87	0.76	0.93	0.91	0.02	—	1.59	3.75	2.13	—	—
K ₂ O	—	—	0.01	—	—	—	0.33	—	—	2.78	11.04	13.71	—	—
Total	99.28	99.00	99.73	99.05	99.30	99.01	99.84	100.32	100.15	97.18	100.33	100.40	—	—
Oxygen	12	12	6	6	6	6	6	5	5	23	8	8	—	—
Si	2.964	2.953	1.953	1.698	1.526	1.529	1.540	0.989	0.981	5.739	2.977	2.991	—	—
Ti	0.022	0.016	0.006	0.026	0.107	0.107	0.231	0.917	0.951	0.484	—	—	—	—
Al	1.935	1.956	0.253	0.603	0.741	0.701	0.560	0.100	0.066	3.006	1.016	1.002	—	—
Cr	—	—	—	—	—	—	0.001	0.001	0.001	—	—	—	—	—
Fe ³⁺	0.103	0.108	—	—	0.049	0.094	—	0.008	0.012	—	0.004	0.009	—	—
Fe ²⁺	0.813	0.036	0.163	0.210	0.171	0.152	0.228	—	—	1.809	—	—	—	—
Mn	0.026	0.095	0.003	—	—	0.006	0.004	0.003	—	0.034	—	—	—	—
Mg	0.642	0.571	0.632	0.488	0.410	0.419	0.451	—	—	1.929	—	—	—	—
Ca	1.486	1.265	0.822	0.919	0.941	0.923	0.893	1.022	1.018	1.913	0.033	0.008	—	—
Na	0.009	0.002	0.167	0.063	0.056	0.069	0.067	0.002	—	0.473	0.330	0.189	—	—
K	—	—	0.001	—	—	—	0.016	—	—	0.543	0.638	0.798	—	—
Cations	8.000	8.000	3.999	4.006	4.000	4.000	3.990	3.041	3.029	15.930	4.997	4.996	—	—

Cpx, clinopyroxene; afsp, alkali feldspar; pcl, porphyroclast; nbl, neoblast. Amphibole was calculated assuming 13 cations, excluding Ca, Na and K. The biotite formula was calculated assuming 11 cations and a $\text{Fe}^{3+}/(\text{Fe}^{2+} + \text{Fe}^{3+})$ ratio of 0.1. Mineral compositions were determined by wavelength-dispersive (WDS) and energy-dispersive (EDS) electron microprobe analysis at Institut für Geowissenschaften, Universität Mainz (CAMECA MICROBEAM) and at Institut für Petrographie und Geochemie, Universität Karlsruhe (CAMECA CAMEBAX SX50). Operating conditions were 15 kV, 10 nA with counting times of 200 s (EDS) or 15 s (WDS, except Na: 30 s). Natural and synthetic standards were used for calibration. PAP (WDS) and ZAF (EDS) correction procedures were applied.

* In garnet.

† In clinopyroxene.

‡ CO₂ not measured.

§ Calculated on the basis of 8 cations.

|| Calculated on the basis of 4 cations.

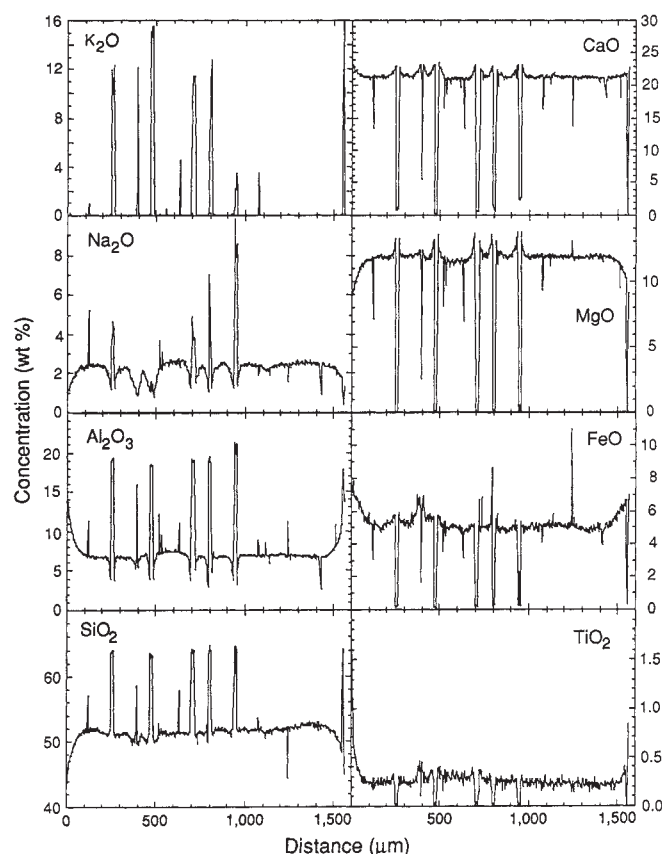


FIG. 3 Compositional profile across cpx porphyroblast (profiles run from left to right along line A-B in Fig. 1). Individual quantitative electron microprobe analyses were taken every 3 μm . Note decreasing Na and Al, but increasing Ca and Mg in cpx host towards alkali feldspar lamellae. Zoning of Al close to the lamellae is the reverse of that observed at the porphyroblast rim. Concentrations of oxides are given in weight %.

alkali feldspar in the cores of primary clinopyroxenes indicate that K_2O contents were considerably higher than the measured values. It is difficult to estimate the original K_2O content of the primary clinopyroxenes because of the large size and irregular distribution of the exsolutions, but values between 0.5 and 1.0 wt% seem most reasonable. The bulk K_2O content of one of the garnet pyroxenite nodules (0.5 wt% K_2O) provides an additional lower limit for K_2O in its cpx, because no K-bearing phases other than the alkali feldspar exsolutions in cpx are present. From studies of diamondiferous eclogite xenoliths²²⁻²⁴, as well as from experiments^{6,7}, it is known that incorporation of potassium into the cpx lattice as a minor component

(>0.1 wt% K_2O) is only possible at pressures above 3–4 GPa. Clinopyroxenes with high K_2O contents and with K-feldspar exsolution lamellae have hitherto been known only from the diamond-bearing ultra-high-pressure rocks from the Kokchetav massif^{4,5} and from diamond-bearing eclogite xenoliths found in some kimberlite pipes²²⁻²⁴.

For this peculiar rock association from Lower Austria, only high-temperature garnet peridotites and garnet pyroxenites yield evidence for comparable ultra-high-pressure conditions (3.5–4 GPa)¹⁴⁻¹⁹. Maximum pressures so far deduced for associated granulites are 1.6–1.8 GPa (ref. 17). Possibly, ultra-high-pressure rocks and granulites were tectonically combined during exhumation from different depths.

The core-to-rim decrease of the grossular component in garnet and the complementary increase of the calcium-tschermaks component in porphyroclastic cpx indicate a pronounced decompression^{25,26}. This kind of zoning can also be observed in Lower Austrian garnet peridotites and garnet pyroxenites^{18,19}. The formation of Al-rich cpx neoblasts also suggests partial re-equilibration of the rock at lower pressures, but still at high temperatures (>1,000 °C).

These 'low-pressure'/high-temperature conditions are most probably similar to the P - T conditions estimated for the early history of the associated garnet-spinel and spinel lherzolites (1.5–2 GPa at 1,150–1,300 °C; refs 18, 19). The similarity in 'low-pressure' equilibration conditions of both calc-silicate marble and peridotites, their close spatial association and the ultra-high P - T origin of the calc-silicate marble imply ascent of these rocks along a gradient similar to the adiabatic upwelling curve for normal-temperature asthenosphere. Low concentrations of some key trace elements (for example, 360 p.p.m. Sr, 160 p.p.m. Ba, 5 p.p.m. Nb) make it unlikely that the calc-silicate marble is of magmatic origin (a former carbonatite). A sediment is a more plausible candidate for the protolith. This sediment must have been subducted into the mantle and equilibrated at ambient mantle conditions. The exhumation path outlined above (near-isothermal decompression at $T > 1,000$ °C) contrasts markedly with the P - T path reconstructed for ultra-high-pressure metamorphics from the western Alps, which formed at much lower temperatures and were cooled considerably during decompression¹. But there are striking similarities to the exhumation path suggested for the Beni Bousera high-temperature peridotite complex. There, diamond pseudomorphs within garnet pyroxenite layers indicate equilibration in the diamond stability field with a subsequent near-isothermal decompression along an upper-mantle adiabat^{27,28}. At least some of these pyroxenite layers represent relics or melts derived from recycled oceanic crust^{28,29}. The calc-silicate marble described here is direct evidence for recycling of sediments into the convecting asthenospheric mantle. It supports geochemical models suggesting that the isotope and trace element composition of some ocean-island basalts require a limited amount of subducted sediment in their mantle source^{9,10}. □

Received 11 May; accepted 15 July 1992.

- Chopin, C., Henry, C. & Michard, A. *Eur. J. Miner.* **3**, 263–291 (1991).
- Smith, D. C. *Nature* **310**, 641–644 (1984).
- Wang, X., Liou, J. G. & Mao, H. K. *Geology* **17**, 1085–1088 (1989).
- Sobolev, N. V. & Shatsky, V. S. *Nature* **343**, 742–746 (1990).
- Jagoutz, E., Shatsky, V. S. & Sobolev, N. V. *EOS* **71**, 1707 (1990).
- Erlank, A. J. & Kushiro, I. *Carnegie Inst. Wash. Yb.* **68**, 233 (1970).
- Shimizu, N. *Earth planet. Sci. Lett.* **11**, 374–380 (1971).
- Laubscher, H. *Geol. Soc. Am. Bull.* **100**, 1313–1328 (1988).
- Hawkesworth, C. J., Norry, M. J., Roddick, J. C. & Vollmer, R. *Nature* **280**, 28–31 (1979).
- White, W. M. & Hofmann, A. W. *Nature* **296**, 821–825 (1982).
- van Breemen, O. *et al. Trans. R. Soc. Edinb., Earth Sci.* **73**, 89–108 (1982).
- Wendt, J. I., Kröner, A., Todt, W., Fiala, J. & Rajlich, P. *Terra Abstr.* **1**, 4 (1989).
- Aftalion, M., Bowes, D. R. & Vrana, S. *Neues Jb. Mineral. Monatshefte* 145–152 (1989).
- Medaris, L. G. & Carswell, D. A. in *Eclogite Facies Rocks* (ed. Carswell, D. A.) 260–290 (Blackie, Glasgow, 1990).
- Carswell, D. A. & Jamveit, B. *Neues Jb. Miner. Abh.* **162**, 69–78 (1990).
- Medaris, L. G., Wang, H. F., Misar, Z. & Jelinek, E. *Lithos* **25**, 189–202 (1990).
- Carswell, D. A. & O'Brien, P. J. *Terra Abstr.* **3**, 93 (1991).

- Becker, H. & Altherr, R. *Geological Workshop Moravian Windows* (eds Schulmann, K., Lobkowicz, M. & Melka, R.) 34 (Geological Survey, Prague, 1991).
- Carswell, D. A. *Eur. J. Miner.* **3**, 323–342 (1991).
- Robinson, P. *Rev. Miner.* **7**, 479 (1980).
- Krogh, E. J. *Contrib. Miner. Petrol.* **99**, 44–48 (1988).
- Reid, A. M., Brown, R. W., Dawson, J. B., Whitfield, G. G. & Siebert, J. C. *Contrib. Miner. Petrol.* **58**, 203–220 (1976).
- Dawson, J. B. & Carswell, D. A. in *Eclogite Facies Rocks* (ed. Carswell, D. A.) 315–349 (Blackie, Glasgow, 1990).
- Harlow, G. E. & Veblen, D. R. *Science* **251**, 652–655 (1991).
- Wood, B. J. & Holloway, J. R. *Geochim. cosmochim. Acta* **48**, 159–176 (1984).
- Gasparik, T. *Contrib. Miner. Petrol.* **87**, 87–97 (1984).
- Pearson, D. G., Davies, G. R., Nixon, P. H. & Milledge, H. J. *Nature* **338**, 60–62 (1989).
- Kornprobst, J., Piboule, M., Roden, M. & Tabit, A. J. *Petrol.* **31**, 717–745 (1990).
- Pearson, D. G., Davies, G. R., Nixon, P. H., Greenwood, P. B. & Matthey, D. P. *Earth planet. Sci. Lett.* **102**, 289–301 (1991).

ACKNOWLEDGEMENTS. We thank G. Brey and K. Mezger for reviewing the manuscript, and H. Frohna-Binder, H.-P. Meyer, B. Schulz-Dobrick, E. Schröder and K. Wacker for technical assistance.

Addition and subtraction by human infants

Karen Wynn

Department of Psychology, University of Arizona, Tucson, Arizona 85721, USA

HUMAN infants can discriminate between different small numbers of items¹⁻⁴, and can determine numerical equivalence across perceptual modalities^{5,6}. This may indicate the possession of true numerical concepts^{1,4-7}. Alternatively, purely perceptual discriminations may underlie these abilities^{8,9}. This debate addresses the nature of subitization, the ability to quantify small numbers of items without conscious counting^{10,11}. Subitization may involve the holistic recognition of canonical perceptual patterns that do not reveal ordinal relationships between the numbers¹², or may instead be an iterative or 'counting' process that specifies these numerical relationships^{4,13}. Here I show that 5-month-old infants can calculate the results of simple arithmetical operations on small numbers of items. This indicates that infants possess true numerical concepts, and suggests that humans are innately endowed with arithmetical abilities. It also suggests that subitization is a process that encodes ordinal information, not a pattern-recognition process yielding non-numerical percepts.

The experiments used a looking-time procedure that has become standard in studies of infant cognition¹⁴⁻¹⁷. Thirty-two infants participated in experiment 1. They were normal, full-term infants with a mean age of 5 months 1 day (range, 4 months 19 days to 5 months 16 days). Infants were divided randomly into two equal groups. Those in the '1+1' group were shown a single item in an empty display area. A small screen then rotated up, hiding the item from view, and the experimenter brought a second identical item into the display area, in clear view of the infant. The experimenter placed the second item out of the infant's sight behind the screen (Fig. 1). Thus, infants could clearly see the nature of the arithmetical operation being performed,

but could not see the result of the operation. The '2-1' group were similarly shown a sequence of events depicting a subtraction of one item from two items (Fig. 1). For both groups of infants, after the above sequence of events was concluded, the screen was rotated downward to reveal either 1 or 2 items in the display case. Infants' looking time to the display was then recorded. Each infant was shown the addition or subtraction 6 times, the result alternating between 1 item and 2 items. Before these test trials, infants were presented with a display containing 1 item and a display containing 2 items and their looking time was recorded, to measure the baseline looking preferences for the two displays.

Infants look longer at unexpected events than expected ones, thus, if they are able to compute the numerical results of these arithmetical operations, they should look longer at the incorrect than at the correct results. The two groups should respond differently to results of 1 and 2 items: the '2-1' group should look longer than the '1+1' group when the result is 2 items than when it is 1 item, which is what is found (Table 1). Pretest trials showed that infants in the two groups did not differ from each other in their baseline looking times to 1 or 2 objects. But in the test trials, infants in the two groups differed significantly—infants in the '1+1' group looked longer at 1, whereas infants in the '2-1' group looked longer at 2. Thus, both groups looked longer at the incorrect than at the correct outcomes (Table 1).

Experiment 2 was a replication of experiment 1 with a smaller number of subjects (sixteen). Their mean age was 4 months 25 days (range, 4 months 18 days to 5 months 5 days). The same pattern of results was obtained; infants in each group looked longer at the incorrect outcome than at the correct outcome (Table 1).

These results show that infants know that an addition or subtraction results in a change in the number of items. But the results are consistent with two distinct hypotheses: (1) that infants are able to compute the precise results of simple additions and subtractions and (2) that infants expect an arithmetical operation to result in a numerical change, but have no expectations about either the size or the direction of the change. They

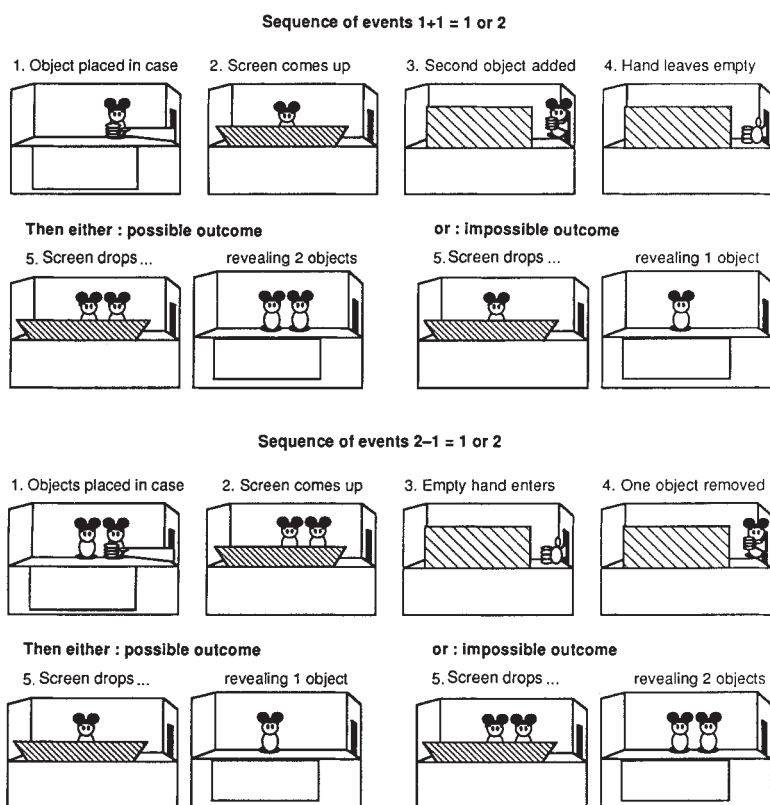


FIG. 1 Sequence of events for '1+1' and '2+1' situations presented in experiments 1 and 2.

METHODS. Trials alternated between a 1-item and a 2-item result, half of the infants received the reverse ordering (1, 2, 1, 2, 1, 2), the remainder receiving the reverse ordering. Infants sat facing the display; parents either stood out of sight behind and not touching the infant, or else gently touched the infant while facing away from the display. The experimenter was hidden behind the display, and manipulated the objects by means of a hidden trap door in the back wall of the display. A hidden observer, unaware of the infant group and of the trial ordering, timed infants' looks to the display. In all experiments, infants were excluded if they became fussy or drowsy during the experiment (16 infants), if their test preference was more than 2.5 standard deviations away from the mean for that group (1 infant), or if they had a pretest preference of more than 10 s for either number (19 infants). The choice of 10 s does not affect the pattern of results (the analyses for experiments 1 and 2 combined give the same pattern even with no cutoff).

TABLE 1 Looking times and preference for 2 items over 1 item

Experiment	Trials	Group	LT(1)*	LT(2)*	P(2)*	d.f.	t	P
1	Pretest	1+1	20.06	20.80	0.74	30	0.649	>0.5
		2-1	17.99	19.61	1.62			
	Test	1+1	13.36	12.80	-0.53	30	2.078	<0.05
		2-1	10.54	13.73	3.19			
2	Pretest	1+1	11.12	10.62	-0.50	14	0.677	>0.5
		2-1	10.35	11.44	1.09			
	Test	1+1	12.08	9.45	-2.65	14	1.795	<0.05
		2-1	10.98	8.05	2.94			
1+2	Pretest	1+1	17.62	18.02	0.41	46	0.873	>0.35
		2-1	15.05	16.47	1.42			
	Test	1+1	13.01	11.89	-1.11	46	2.73	<0.005
		2-1	9.59	12.67	3.09			

Statistical significance was determined by between-group *t*-tests on infants' P(2) values. Probability values are 2-tailed for pretest comparisons, 1-tailed for test comparisons. In experiment 1, a trial concluded when an infant looked away for 2 consecutive seconds after looking at the display for at least 4 cumulative seconds, or had looked for 30 cumulative seconds. Experiment 2; same criteria, except that minimum cumulative looking time was only 2 s. The shorter mean looking times in experiment 2 are probably due to this procedural change. Times are lower in test than pretest trials because infants' looks decrease during the experiment as they become more familiar with the display. Experiment 2, 6 infants in the 1+1 group, 10 infants in the 2-1 group.

* P(2)=LT(2)-LT(1); where P(2), preference for 2; LT(1) and LT(2) are the mean looking times to 1 and 2 items (in seconds).

may simply expect that adding an item to an item will result in some number other than 1; and that subtracting an item from 2 items will result in some number other than 2. To determine whether infants are able to compute the precise results of simple arithmetical operations, I conducted a third experiment.

Experiment 3 tested 16 infants with a mean age of 4 months 18 days (range, 4 months 4 days to 5 months 4 days). Infants were shown a '1+1' addition as before, except that the final number of objects revealed behind the screen was either 2 or 3. In both cases, the result is numerically different from the initial number of items. If infants are computing the exact numerical result of the addition, they would be expected to look longer at the result of 3 items than of 2 items. This pattern was indeed observed (Table 2); infants significantly preferred 3 in the test trials, but not the pretest trials, showing that they were surprised when the addition appeared to result in 3 items. The results from the three experiments support the claim⁷ that 5-month-old human infants are able to calculate the precise results of simple arithmetical operations.

There is an alternative explanation for infants' success in these experiments. Infants may be calculating the results of the addition and subtraction, not of a discrete number of items, but of a continuous amount of physical substance; infants may possess an ability to measure and operate on continuous quantities. But there are reasons to prefer the hypothesis that it is the number of items, not amount of substance, that infants are computing. It has been shown that infants are sensitive to small numerical changes¹⁻⁴, but there is no evidence of a sensitivity to small differences in amount of physical matter. Infants are predisposed to interpret the physical world as composed of discrete, individual entities when perceiving spatial layouts^{14,15}, and they represent the precise spatial locations and trajectories of individual objects relative to each other^{16,17}. Thus, the notion of 'individual entity' plays a prominent role in infants' conceptualization and representation of the physical world, and they

TABLE 2 Looking times and preferences for 3 items over 2

Condition	LT(2)*	LT(3)*	P(3)*	d.f.	t	P
Pretest	14.16	13.87	-0.29	15	-0.224	>0.5
Test	9.96	11.89	1.92	15	2.044	<0.03

Statistical significance was determined by *t*-tests comparing infants' P(3) values to the null hypothesis of no preference. Probability value for pretest comparison is 2-tailed; that for test comparison is 1-tailed. As in experiments 1 and 2, infants were excluded if they showed more than a 10-second pretest preference for one of the numbers; the pattern of results remains the same when these infants are included in the analyses. Experiment 3 used the same criterion for end-of-trial as that used in experiment 2.

* P(3)=LT(3)-LT(2), where P(3), preference for 3; LT(3) and LT(2) are the mean looking times to 3 and 2 items (in seconds).

have abilities that allow them to track distinct entities over time and space. This, together with infants' sensitivity to small numerical differences in collections of items, lends independent support to the hypothesis that infants possess a mechanism for quantifying collections of discrete entities. The most plausible explanation for the findings presented here is that infants can compute the results of simple arithmetical operations.

In sum, infants possess true numerical concepts—they have access to the ordering of and numerical relationships between small numbers, and can manipulate these concepts in numerically meaningful ways. This in turn indicates that the mental process giving rise to these concepts yields true numerical outputs that encode numerical relationships, not holistic percepts derived from a pattern-recognition process. The existence of these arithmetical abilities so early in infancy suggests that humans innately possess the capacity to perform simple arithmetical calculations, which may provide the foundations for the development of further arithmetical knowledge^{7,18}. □

Received 20 May; accepted 16 July 1992.

- Starkey, P. & Cooper, R. G. *Science* **210**, 1033-1035 (1980).
- Strauss, M. S. & Curtis, L. E. *Child Dev.* **52**, 1146-1152 (1981).
- Antell, S. & Keating, D. P. *Child Dev.* **54**, 695-701 (1983).
- van Loosbroek, E. & Smitsman, A. W. *Dev. Psychol.* **26**, 916-922 (1990).
- Starkey, P., Spelke, E. S. & Gelman, R. *Science* **222**, 179-181 (1983).
- Starkey, P., Spelke, E. S. & Gelman, R. *Cognition* **36**, 97-127 (1990).
- Wynn, K. *Mind Lang.* (in the press).
- Davis, H., Albert, M. & Barron, R. W. *Science* **228**, 1222 (1985).
- Cooper, R. G. in *Origins of Cognitive Skills* 157-192 (ed. Sophian, C.) (Erlbaum, Hillsdale, New Jersey, 1984).
- Chi, M. T. H. & Klahr, D. *J. exp. Child Psychol.* **19**, 434-439 (1975).
- Silverman, I. W. & Rose, A. P. *Dev. Psychol.* **16**, 539-540 (1980).

- Mandler, G. & Shebo, B. J. *J. exp. Psychol. Gen.* **11**, 1-22 (1982).
- Gallistel, C. R. *The Organization of Learning* 343-348 (MIT, Cambridge, Massachusetts, 1990).
- Spelke, E. S. in *Perceptual Development in Infancy: Minnesota Symposia on Child Psychology* Vol. 20 (ed. Yonas, A.) 197-234 (Erlbaum, Hillsdale, New Jersey, 1988).
- Spelke, E. S. *Cog. Science* **14**, 29-56 (1990).
- Baillargeon, R. *Cognition* **38**, 13-42 (1991).
- Spelke, E. S. in *Perceptual Development in Infancy: Minnesota Symposia on Child Psychology* Vol. 20 (ed. Yonas, A.) 197-234 (Erlbaum, Hillsdale, New Jersey, 1988).
- Baillargeon, R. & DeVos, J. *Child Dev.* **62**, 1227-1246 (1991).
- Wynn, K. *Cog. Psychol.* **24**, 220-251 (1992).

ACKNOWLEDGEMENTS. I thank E. S. Spelke and R. Baillargeon for advice and suggestions; F. Bedford, B. Bloom, M. Peterson, and R. Rosser for suggestions and comments; and H. Campbell, K. Fohr, C. Mannering, and T. Wilcox for help in the testing of infants. This was supported by the NIH (Institutional BRSG grant) and the University of Arizona (SBSRI grant).

A test of reciprocal X-Y interactions as a cause of hybrid sterility in *Drosophila*

Norman A. Johnson*, Daniel E. Perez†, Eric L. Cabot†, Hope Hollocher† & Chung-I Wu†‡

* Department of Biology, University of Rochester, Rochester, New York 14627, USA

† Department of Ecology & Evolution, University of Chicago, Chicago, Illinois 60637, USA

ELUCIDATION of the nature of the gene interactions that underly the sterility of interspecific hybrids is important in evolutionary biology^{1,2}. The interactions between the heterospecific X and Y (or Z and W) chromosomes are often used as an explanation for two reasons. First, the fertility of the hybrids of the heterogametic sex is much more often affected than that of the homogametic sex (Haldane's rule³) and X-Y interactions are specific to the heterogametic sex. Second, sex chromosomes, especially the X chromosome, are often considered to be of special importance in determining the fertility of hybrids^{1,2,4}. X-Y interactions have been addressed in studies of males with a heterospecific Y chromosome in a mixed genetic background⁵⁻⁸. A more stringent test of the X-Y interaction model requires each X chromosome sterility factor to be tested separately for its interaction with the Y chromosome in a homogeneous background of the pure species. Here we report such a test of the X-Y interaction model and conclude that X-Y interactions should not be assumed to be the only or even the most common cause of hybrid sterility.

Crosses between *Drosophila simulans* and its sibling species *D. sechellia* result in fertile F₁ females and sterile F₁ males, and thus follow Haldane's rule. At least three X-linked factors that cause male sterility when introgressed from *D. sechellia* into *D. simulans* have been mapped by linkage analysis⁹. We used an approach that relies on interspecific DNA variations as markers to construct a fine-scale physical map of X-linked sterility factors between *D. simulans* and either of its two sibling species, *D. mauritiana* or *D. sechellia*¹⁰. We constructed a stock with a Y chromosome introgressed from *D. sechellia* into a nearly pure *D. simulans* background (type A males in Fig. 1; for details of the construction of type A males, see ref. 11). The identity of the Y chromosome was further confirmed by probing genomic DNA from both sexes with the Stellate sequence¹², which hybridizes predominantly to the Y chromosome (Fig. 2). The fertility of type A males indicates the absence of sterility interactions between Y_{sec} (the Y chromosome of *D. sechellia*) and the genome of *D. simulans*. In fact, only after repeated mating to several females do type A males produce fewer offspring than the pure *D. simulans* males (N.A.J., unpublished results). The fertility rules out interactions between Y_{sec} and genes on X_{sim} (the X chromosome of *D. simulans*) but provides no information on the reciprocal interactions between X_{sec} and Y_{sim} because such gene interactions are often asymmetric for hybrid sterility^{13,14}.

In a parallel experiment, we attempted to introgress the Y chromosome of *D. simulans* into *D. sechellia* but were unsuccessful in producing fertile males with Y_{sim}. The lack of success suggests sterility interactions between Y_{sim} and *D. sechellia* genes ($P < 0.01$ if Y_{sim} has no sterility effect; see the legend to Fig. 2). In other words, the reciprocal genotype of type A (with blank and solid bars reversed) appears sterile. In an independent study, Zeng and Singh (personal communication) succeeded in introgressing Y_{sec} into *D. simulans* by a different method, which also failed to yield *D. sechellia* males with an introgressed Y_{sim}. Because X_{sec} interacts with the *D. simulans* background (the sterile type B males of Fig. 1) and Y_{sim} apparently interacts with

D. sechellia genes because it was non-introgressable, it is natural to formulate the null hypothesis that X_{sec} (or some segments of it) interact with Y_{sim}. An alternative hypothesis is that X_{sec} interacts with the autosomes of *D. simulans*¹ and Y_{sim} with the autosomes of *D. sechellia*¹⁵ but X_{sec} and Y_{sim} do not interact with each other.

Our test of the null hypothesis of X_{sec}-Y_{sim} interactions consists of two steps. (1) Three different X-linked regions were individually introgressed from *D. sechellia* into *D. simulans* by repeatedly backcrossing hybrid females to *D. simulans* males for more than 15 generations¹⁰. Each introgression contains a factor(s) that completely sterilizes the males. Sterility must be the result of interactions between the small introgressed segment and the *D. simulans* genome because the introgressed segment itself does not cause sterility in its own *D. sechellia* background. An example of such an introgression genotype is shown in Fig. 1 (type B). (2) We then took advantage of the fact that the Y_{sec} chromosome is fully introgressable into *D. simulans* and replaced the Y_{sim} chromosome of the type B sterile males with the Y chromosome from *D. sechellia* (type C of Fig. 1). If sterility is caused by X-Y interactions, co-introgression of Y_{sec} with each X_{sec} sterility factor should lead to fertility rescue, or at least to a more advanced spermatogenic development.

The results are summarized in Table 1 for each of the three regions of the X chromosome, marked with the visible markers, *y* (yellow, 1B on the polytene chromosome map), *v* (vermillion, 10A) or *f* (forked, 15F), respectively. Co-introgression of Y_{sec} has no effect on the sterility caused by any of the three introgressions. The criteria for sterility are that no progeny are produced in test crosses and no motile sperm can be detected in the male's seminal vesicle. (Note that Y_{sec} alone does not

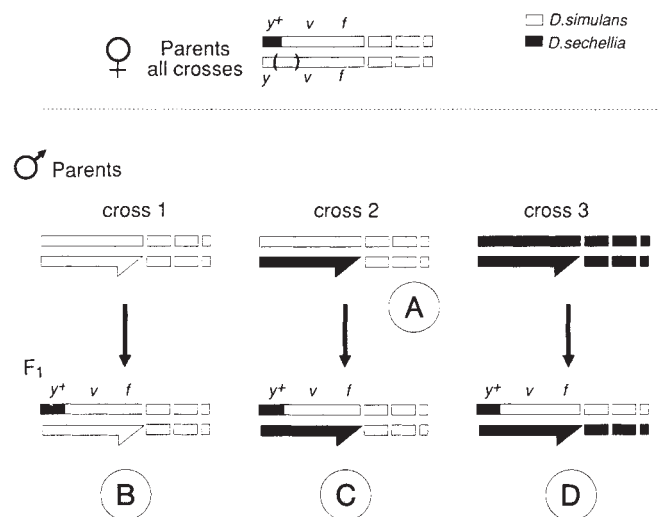


FIG. 1 Crosses generating males of genotypes B, C and D. These males are all sterile (their spermatogenic phenotypes given in Table 1), whereas type A males have normal fertility. The genotype of the female parent is the same for all three crosses and is given at the top. These females have been backcrossed to *D. simulans* males for at least 15 generations¹⁰. The X chromosome, marked with *y*, *v* and *f*, is drawn to a larger scale. The parentheses represent the breakpoints of an inversion at 2B and 8B. This inversion ensures that almost all *y*⁺-bearing chromosomes in F₁ males carry the same introgression without undergoing crossing-over in that region¹⁰. Type A males, the male parent in cross 2, were constructed as described in ref. 11, crosses 1 and 2 were repeated for two other introgressions marked with *y*⁺ and *f*⁺, respectively. There are no inversions to prevent crossover in these two regions. For the *f* region, the *D. simulans* chromosome is marked with *g* (12B), *f* (15F) and *Bx* (17A) markers, thus ensuring the introgression spanning 12B and 17A would not be eroded by recombination. For the *v* region, we relied only on male sterility itself to ensure that the sterility factor is still present. Periodically, we used a single female parent to continue the cross in order to purge the stocks of heterogeneity in the size of introgression due to recombinations.

‡ To whom correspondence should be addressed.

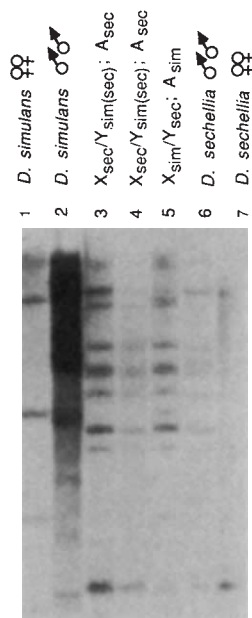


FIG. 2 Molecular identification of the introgressed Y chromosome. We used the Stellate sequence (plasmid pSX1.3 of ref. 12) to probe genomic DNA of *D. simulans* females (lane 1), *D. simulans* males (lane 2), *D. sechellia* males (lane 6), *D. sechellia* females (lane 7) and the three introgression lines described below. (The signals shown are largely Y-specific but some X-specific bands are also identifiable. These Y-specific signals are species-specific.) The X_{sim}/Y_{sec} males (lane 5) are *D. simulans* flies with an introgressed Y_{sec} , as described in ref. 11. The identity of the Y chromosome is confirmed to be of the *D. sechellia* origin. The $X_{sec}/Y_{sim(sec)}$ males (lanes 3 and 4; two independent lines) were expected to carry Y_{sim} but in fact carry Y_{sec} . These males were obtained by crossing $C(1)yw/Y_{sim}; (1/2A_{sim}, 1/2A_{sec})/A_{sim}$ females to *D. sechellia* males. $C(1)yw$ is the attached-X chromosomes of *D. simulans* and A denotes autosomes. These $C(1)yw$ females were from the F_2 generation of the cross of ref. 11. The expected genotype of their male progeny (F_3) is $X_{sec}/Y_{sim}; (3/4A_{sim}, 1/4A_{sec})/A_{sec}$. But because of possible nondisjunction between $C(1)yw$ and the Y chromosome in previous generations, a small fraction of the F_2 females may be $C(1)yw/Y_{sim}/Y_{sec}$ and so half of their F_3 sons could be $X_{sec}/Y_{sec}; (3/4A_{sim}, 1/4A_{sec})/A_{sec}$. Among 3,800 F_3 male progeny, possibly from such a mixture of F_2 females, only two were fertile. (The high incidence of sterility is due to interactions between the autosomes of *D. simulans* and either the X or the autosomes of *D. sechellia*.) Each of these two males were backcrossed to *D. sechellia* virgin females for seven generations to establish a pure line and then their Y chromosomes were examined for the Stellate genotype. The fact that both fertile males were X_{sec}/Y_{sec} , instead of the expected X_{sec}/Y_{sim} genotype, suggests that Y_{sim} causes male sterility when introgressed into the genome of *D. sechellia*. A statistical argument can be made as follows. (1) In F_3 , X_{sec}/Y_{sim} males are estimated to outnumber X_{sec}/Y_{sec} males by 25-fold. (We obtained this estimate by counting the metaphase Y chromosomes in the neuroblasts of $C(1)yw$ females and found one out of the 25 females examined to have two Y chromosomes. The rate of *de novo* nondisjunction in these females is negligible because wild-type male progeny of $C(1)yw$ females are often 100% fertile.) (2) Because the background sterility interactions in F_3 males (that is between X and autosomes or between autosomes) are the same for both types, the probability of recovering a X_{sec}/Y_{sec} male should only be half of $1/25 (=0.02)$, if the introgression of Y_{sim} does not cause sterility. Under that assumption, the probability that the only two fertile males recovered are both X_{sec}/Y_{sec} should be much less than $0.01 (>0.02^2)$. The results thus invalidate an earlier statement of ours in ref. 11 claiming the fertility of $X_{sec}/Y_{sim}; A_{sec}/A_{sec}$ males. The observations provide a strong argument for molecular identification of introgressed materials.

METHODS. Genomic DNA was digested with *Xba*I, electrophoresed in a 0.8% agarose gel and blotted onto a Hybond-N membrane. Hybridization was done at 60 °C in 6×SSPE, Blotto and 1% SDS overnight. The membrane was washed for 15 min in 4×SSC and 0.2% SDS at room temperature, then 4×SSC and 0.1% SDS for 30 min at 37 °C and finally in 2×SSC and 0.1% SDS for 30 min at 37 °C.

TABLE 1 Fertility and spermatogenic development in males with genotypes shown in Fig. 1

	Genotype B	Genotype C	Genotype D
<i>y</i> region	Sterile Small deformed testes $3.40 \pm 1.43^* (10)$ Much reduced number of spermatocytes	Sterile Small deformed testes $3.93 \pm 0.96^* (36)$ Much reduced number of spermatocytes	Sterile Normally shaped testes $5.85 \pm 0.92 (22)$ Number and size of spermatocytes and spermatids normal†
<i>f</i> region	Sterile Normally shaped testes $3.93 \pm 0.73 (18)$ Smaller spermatocytes, number reduced	Sterile Normally shaped testes $4.90 \pm 1.24 (10)$ Smaller spermatocytes, number reduced	No data
<i>v</i> region	Sterile	Sterile	No data
Control (fertile males)	$7.38 \pm 0.58 (20)$		

These are the results of crosses in Fig. 1 repeated for three different regions of the X chromosome (see text). The numbers are the length of the testes $\pm$ s.e., representing the maximal dimension of the testes in their normal coiled state. The units are in 0.1 mm. The numbers in parentheses are the number of testes, measured and observed cytologically. The control value is from the males of the *y*, *v*, *f* stock of *D. simulans*. The term spermatocyte refers to mature primary spermatocytes ready to enter meiosis. All descriptions are relative to the fertile males.

* The length of the deformed testes are overestimates of their actual sizes because these testes are not coiled and are much thinner (Fig. 3); hence, in comparison with normally shaped testes of equal length, their areas seen under a microscope are less than half of the normal ones.

† Occasionally, low numbers of motile sperm were observed in these males.

cause any spermatogenic defect in the *D. simulans* genome.) Because these criteria do not take into account the possibility of partial rescue of spermatogenic development in type C males, which remain sterile, we also compared the testes and spermatocytes in males of various genotypes (Table 1).

With the introgressed region marked with y^+ (wild type of yellow), sterility is characterized by very small testes as shown in Fig. 3b. The total volume is on average less than 20% of the normal ones. These testes do not have the normal coiled appearance (thus making its length measurement an underestimate of the severity of the defects). The mature primary spermatocytes, which are ready for meiosis, in the normal and type B males are shown in Fig. 3c and d, respectively. In general, there are very few mature primary spermatocytes in the testes of the sterile type B males. To show the difference, we selected a sterile sample that did have some spermatocytes (Fig. 3d), in contrast to one from a typical fertile male (Fig. 3c). The defect is clearly pre-meiotic and is much more severe than that in sterile F_1 males, which have normal-looking meiocytes (D.E.P., unpublished results). The important point is that, in type C males, neither the testicular morphology nor the spermatogenic development seems different from those of type B males. The testes of type C males are on average slightly larger, but the difference is not statistically significant.

Spermatogenic defects associated with the introgression of the forked region are less severe than those of the yellow region. The testes in type B males are shorter than normal but have the normal appearance and, hence, a volume closer to that of normal males. Mature spermatocytes are abundant, although smaller and not as numerous as normal males. As in the case of the yellow region, cytological observations on type C males are indistinguishable from those on type B males; thus the Y replacement has no observable effect on the spermatogenic development arrested by the f^+ -linked factor.

Ideally, the introgression in type B males should contain only one sterility factor. Preliminary results from molecular mapping are indeed suggestive, although not conclusive, of only one major factor in the *y* and *f* regions from *D. sechellia*¹⁰. If there are two independent sterility factors on the introgressed segment, our sterility test can only rule out that both factors interact with Y_{sim} . (For this reason, previous tests using whole X introgression, for example ref. 8, only ruled out that all X-linked factors interact with Y.) But the absence of rescue at the cytological level enables us to make a stronger inference: if there are two factors on the introgressed segment and one of them interacts with Y_{sim} , the replacement of Y_{sec} might, in some instances, improve the spermatogenic development. Because such a phenomenon was not observed in either of the two regions analysed, the possibility for undetected X-Y interactions is reduced. A more direct line of evidence will be the confirmation of a single factor by molecular means, which is still in progress.

The question is: with what do the introgressed X-linked factors interact to cause male sterility? Because X-autosome interactions seem to be the most likely candidate, we constructed type D males as in Fig. 1. This is not a direct test as the autosomes introduced from *D. sechellia* for rescue create sterility on their own (by interacting with X or autosomes of *D. simulans*). Nevertheless, type D males with the y^+ -region introgression have a closer to normal spermatogenic development than the corresponding type C males (Table 1). Although the mature spermatocytes and the spermatids (that is, the immediate products of meiosis) are usually not observed in type B or C males, they are readily observable in type D males. The difference between type C and D males is that type D males carry a copy of autosomes from *D. sechellia* (A_{sec}), which may provide the missing X_{sec} - A_{sec} interactions required for testicular development. The observation is consistent with the X-autosome interaction model proposed 50 years ago^{1,2}.

In the cross between *D. simulans* females and *D. sechellia* males, sterility does not result from the interaction between Y_{sec} and *D. simulans* genes, including X_{sim} , because it is possible to introgress Y_{sec} into an otherwise pure *D. simulans* background. In the reciprocal cross, several X_{sec} genes (the three sterility factors marked by *y*, *v* and *f*) interact with the *D. simulans*

background, whereas Y_{sim} apparently interacts with *D. sechellia* genes (hence the non-introgressability of Y_{sim} into the *D. sechellia* background). Replacing Y_{sim} with Y_{sec} does not lessen the sterility effect of any of the three X_{sec} genes tested, implying that X-autosome and Y-autosome interactions are causing sterility rather than X-Y interactions. In *Drosophila*, the Y chromosome either has no effect on hybrid sterility^{7,8} or, when it does, has not been shown to interact with the heterospecific X in a homogeneous genetic background. Inferences on the genic interactions of hybrid sterility from heterogeneous complex genotypes, such as the backcross F_2 hybrids, can be divergent even from essentially the same results; for example, between those in refs 1 and 6. In another study where Y causes sterility in a homogeneous background (A. C. Pantazidis, V. K. Galanopoulos and E. Zouros, manuscript in preparation), its interaction was clearly demonstrated with an autosomal factor by means of fertility rescue. Although it has often been accepted^{16,17}, evidence for the X-Y interaction model of hybrid sterility has remained inconclusive. It is thus prudent not to emphasize X-Y interactions at the expense of other plausible models such as sex chromosome-autosome interactions. □

Received 28 April; accepted 2 July 1992.

1. Dobzhansky, T. *Genetics* **21**, 113-136 (1936).
2. Muller, H. J. in *The New Systematics* (ed. Huxley, J. S.) 185-268 (Clarendon, Oxford, 1940).
3. Haldane, J. B. S. *J. Genet.* **12**, 101-109 (1922).
4. Coyne, J. A. & Orr, H. A. in *Speciation and its Consequences* (eds Otte, D. & Endler, J. A.) 181-207 (Sinauer, Sunderland, Massachusetts, 1989).
5. Coyne, J. A. *Nature* **314**, 736-738 (1985).
6. Orr, H. A. *Genetics* **116**, 555-563 (1987).
7. Hennig, W. *Adv. Enz. Regul.* **15**, 363-371 (1977).
8. Orr, H. A. *Evolution* **43**, 180-189 (1989).
9. Coyne, J. A. & Charlesworth, B. *Heredity* **62**, 97-106 (1989).
10. Wu, C.-I. et al. in *Molecular Paleo-population Biology* (eds Takahata, N. & Clark, A. G.) (Springer, Berlin, in the press).
11. Johnson, N. A. & Wu, C.-I. *Genetics* **130**, 507-511 (1992).
12. Livak, K. *Genetics* **107**, 611-634 (1984).
13. Oka, H. *Genetics* **77**, 521-534 (1974).
14. Wu, C.-I. & Beckenbach, A. T. *Genetics* **105**, 71-86 (1983).
15. Pantazidis, A. C. & Zouros, E. *Heredity* **60**, 299-304 (1988).
16. Jones, J. S. & Barton, N. *Nature* **314**, 668-669 (1985).
17. Clark, A. *Genetics* **115**, 143-151 (1987).

ACKNOWLEDGEMENTS. We thank A. Davis, M. Palopoli, J. Coyne and A. Clark for helpful comments. The *Stellate* clone was provided by A. Clark. The work is supported by an NIH training grant to N.A.J., an Alfred P. Sloan Foundation Fellowship to E.L.C., an NIH grant, an RCDA award and an NDSS Fellowship to C.I.W.

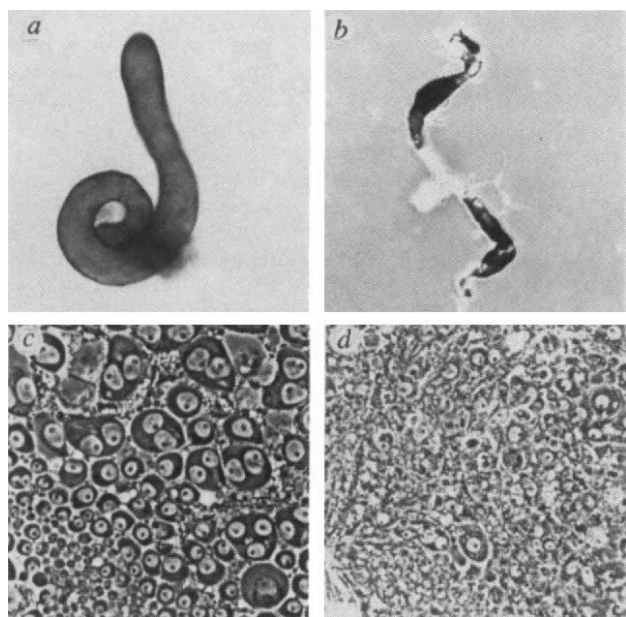


FIG. 3 A severe phenotype of male sterility resulting from the introgression of an X-linked factor. *a, c*, Whole testis and primary spermatocytes in pure *D. simulans* males. *b, d*, A pair of testes and primary spermatocytes in *D. simulans* males whose yellow-locus region is introgressed from *D. sechellia*.

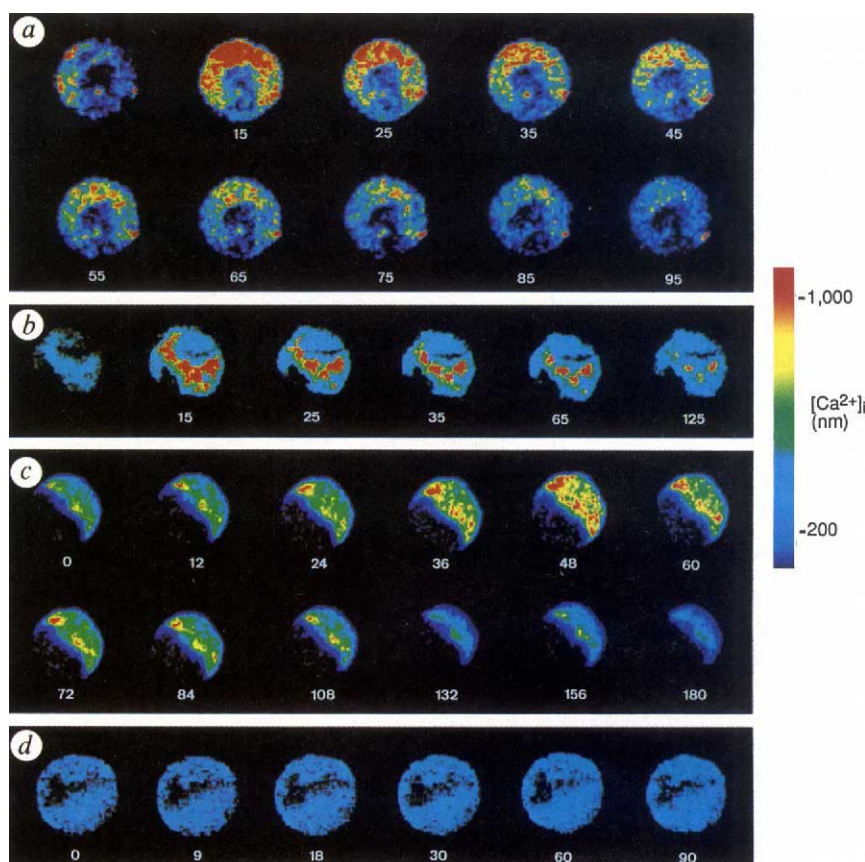
Cytosolic free calcium mediates red light-induced photomorphogenesis

P. S. Shacklock, N. D. Read & A. J. Trewavas

Molecular Signalling Group, Institute of Cell and Molecular Biology, University of Edinburgh, Rutherford Building, Edinburgh EH9 3JH, UK

LIGHT is a primary environmental signal regulating plant growth and form¹. The receptor phytochrome responds to red light and induces changes in membrane properties and gene expression^{2,3} through an unknown transduction pathway which is the subject of intense study. Etiolated wheat leaf protoplasts swell in response to red light⁴, by a mechanism believed to be similar to that involved in phytochrome-regulated leaf growth and unrolling⁵⁻⁷. Here we report that this physiological response is preceded by a transient increase in free calcium in the cytosol, followed by a decrease to below resting level. The timing of this response varied between protoplasts. Cytosolic calcium transients induced by photolytic release of calcium or inositol 1,4,5-trisphosphate from chemically caged forms⁸⁻¹³ inside these protoplasts resulted in red light-induced increases in protoplast volume being mimicked. Our results support the hypothesis that phytochrome-mediated signals

FIG. 1 Visualization of $[Ca^{2+}]_i$ dynamics induced by the photo-release of Ca^{2+} using caged probes, and by red light. $[Ca^{2+}]_i$ was imaged in Fluo-3 loaded etiolated wheat protoplasts using an MRC-600 laser scanning confocal microscope (BioRad) equipped with a transmission detector and a 25 mW argon ion laser providing fluorescence excitation at 488 nm¹¹. Confocal images were collected at 3 s intervals. Changes in $[Ca^{2+}]_i$ were induced by: (a) 25 s ultraviolet photolysis of incorporated caged Ca^{2+} (Nitr-5); (b) 25 s ultraviolet photolysis of incorporated caged $InsP_3$; (c) 3 min red light; and (d) 2 min red light followed by 2 min far red light. Protoplasts used in treatments c and d were not loaded with caged probes. Protoplasts were isolated as in ref. 6 from 9–12-day-old, dark-grown, etiolated wheat leaves (var. Mercia). Protoplast isolation was performed under dim green light (primary green filter, Lee colortran, Wembley, UK). Fluo-3 (Molecular Probes, Eugene, OR, USA) on its own or with caged Ca^{2+} (Nitr-5/AM from calbiochem, Nottingham, UK) was ester-loaded by incubation at 32 °C for 2 h in an incubation medium containing 75 μ M Fluo-3/AM (and 75 μ M Nitr-5, where appropriate), 0.5 M sorbitol, 10 mM MES, 5 mM $CaCl_2$ at pH 6.7 Fluo-3 0.75 μ M, (tetrasodium salt from Molecular Probes) and 50 μ M caged $InsP_3$ (Calbiochem) were loaded by electroporation (1,200 V cm⁻¹ field strength, 0.25 μ F capacitance and three pulses delivered at 5-s intervals). Protoplasts were washed twice in incubation medium before observation. The caged probes were photoactivated with ultraviolet light provided by the microscope 100 W Hg vapour lamp and an interference filter (340 nm, 10 nm half bandwidth (HBW)). Single protoplasts were irradiated with red light (660 nm, 10 nm HBW; fluence rate 100 μ mol m⁻² s⁻¹) and far red light (730 nm, 10 nm HBW; fluence rate 110 μ mol m⁻² s⁻¹) provided by the microscope 100 W Hg vapour lamp and interference filters. The images are representative of the following numbers of replicate experiments: a, *n*=12, b, *n*=13, c, *n*=23 and d, *n*=8. Times are in seconds after illumination. The first images in a and b were obtained before photoactivation of the caged probes. Calibration of Fluo-3 was as described in ref. 10.



are transduced through calcium and help to resolve a long-standing issue in plant physiology.

Wheat leaf protoplasts were loaded with Fluo-3 and caged Ca^{2+} or caged inositol 1,4,5-trisphosphate ($InsP_3$)^{8–10}, and cytosolic free Ca^{2+} ($[Ca^{2+}]_i$) imaged using laser scanning confocal microscopy (LSCM)¹¹. The caged probes were photoactivated with ultraviolet light. $[Ca^{2+}]_i$ was elevated to $>1 \mu$ M in localized regions of the cytosol and returned to the original resting level within 1 min (Fig. 1a and b). Once Ca^{2+} was released, protoplast volume increased by $>20\%$ within 10 min (Fig. 2a). Thus a transient rise in $[Ca^{2+}]_i$ lasting no more than 1 min is sufficient to elicit the prolonged physiological response of a change in cell volume. These data also show that $InsP_3$ can mobilize $[Ca^{2+}]_i$ in etiolated wheat leaf protoplasts, as previously reported for other plant cells^{10,12}. In control experiments ultraviolet light alone had no detectable effect on either protoplast $[Ca^{2+}]_i$ or volume (data not shown).

LSCM of Fluo-3 was also used to examine red light-induced $[Ca^{2+}]_i$ changes. In many cases, $[Ca^{2+}]_i$ was transiently elevated, peaking at about 1μ M in localized regions of the cytosol (Fig. 1c). Protoplast swelling was initiated just before or at the end of the illumination period and protoplast volume increased for at least 5–10 min (Fig. 2b). After the transient increase, $[Ca^{2+}]_i$ then continued to decline below the initial resting level and remained at this very low concentration for the period of observation (8–10 min) in all protoplasts examined.

To demonstrate putative phytochrome involvement in the mediation of these red light-induced responses, protoplasts (*n* = 10) were irradiated with far-red light immediately after the red light treatment. In none of these protoplasts did $[Ca^{2+}]_i$ levels

change during the period of observation (Fig. 1d). The protoplasts observed showed a slight shrinkage (Fig. 2b)⁶. Thus the red light-induced transient and subsequent decrease to below resting level is abolished when red light irradiation is followed by far-red light. We deduce, therefore, that the effects of red light on $[Ca^{2+}]_i$ after red light irradiation and the $[Ca^{2+}]_i$ elevation, if observed. It is possible that differences in cell type may explain this observed heterogeneity as specific cell types were not selected from the protoplast population obtained from the whole leaf. Because these protoplasts are highly vacuolate and the $[Ca^{2+}]_i$ elevations localized, average Fluo-3 fluorescence measurements from the whole protoplasts show an apparently diminished $[Ca^{2+}]_i$ response compared with the results obtained from imaging (Fig. 1a–c). To determine the magnitude of the $[Ca^{2+}]_i$ transient seen in Fig. 3a, we performed ratio imaging using the dual emission dye Indo-1 (refs 13, 14). The $[Ca^{2+}]_i$ rose from initial values of 250 ± 20 nm to a maximum of 900 nm and then fell to $100 \text{ nM} \pm 25$ 2–3 min later (*n* = 10).

Variations in the timing of red light-induced changes in $[Ca^{2+}]_i$ were observed between protoplasts (Fig. 3). We have classified this continuum of variation into three basic patterns. In one group of protoplasts (*n* = 11), the $[Ca^{2+}]_i$ transient began after the end of red light illumination (Figs 1c and 3a). In the second group (*n* = 8), the $[Ca^{2+}]_i$ elevation was initiated before the end of illumination and was above resting level when illumination ceased (Fig. 3b). The final group (*n* = 5) showed no apparent increase in $[Ca^{2+}]_i$. All three groups exhibited a rapid decline in $[Ca^{2+}]_i$ after red light irradiation and the $[Ca^{2+}]_i$ elevation, if observed. It is possible that differences in cell type may explain this observed heterogeneity as specific cell types were not selected from the protoplast population obtained from the whole leaf. Because these protoplasts are highly vacuolate and the $[Ca^{2+}]_i$ elevations localized, average Fluo-3 fluorescence measurements from the whole protoplasts show an apparently diminished $[Ca^{2+}]_i$ response compared with the results obtained from imaging (Fig. 1a–c). To determine the magnitude of the $[Ca^{2+}]_i$ transient seen in Fig. 3a, we performed ratio imaging using the dual emission dye Indo-1 (refs 13, 14). The $[Ca^{2+}]_i$ rose from initial values of 250 ± 20 nm to a maximum of 900 nm and then fell to $100 \text{ nM} \pm 25$ 2–3 min later (*n* = 10).

We considered the possibility that the three patterns of red light-induced $[Ca^{2+}]_i$ responses (Fig. 3) merely reflected variation in the timing of the initiation of the $[Ca^{2+}]_i$ elevation¹⁵. To test this, we incubated Indo-1-loaded protoplasts in 5 mM Mn^{2+} and 1 mM Ca^{2+} and irradiated with red light. Mn^{2+} acts as a surrogate for Ca^{2+} and can thus enter a cell through open Ca^{2+} channels. Because the fluorescence of Ca^{2+} dyes is quenched by Mn^{2+} , this effect can be used to monitor the uptake of Mn^{2+} into dye-loaded cells and thus Ca^{2+} channel activity in the plasma membrane^{15,16}. For this purpose, we examined the quenching of red light-irradiated protoplasts loaded with Indo-1 and found that the dye was quenched in all protoplasts within 4 min of the start of irradiation. This quenching occurred much more rapidly than dye photobleaching by the ultraviolet light source. Our results therefore support the hypothesis that red light induces Ca^{2+} entry into all protoplasts.

Red light-induced Ca^{2+} efflux from coleoptile sections has been measured using $^{45}Ca^{2+}$ (ref. 17). If, as observed here, the initiation of the short-lived Ca^{2+} elevation is asynchronous in individual coleoptile cells, and this elevation is followed by a longer-lived Ca^{2+} efflux, then it is likely that the efflux would dominate average measurements of $[Ca^{2+}]_i$ in cell populations. Our data emphasize the importance of making single-cell measurements if asynchronous $[Ca^{2+}]_i$ transients are to be clearly detected in cell populations. A similar conclusion has

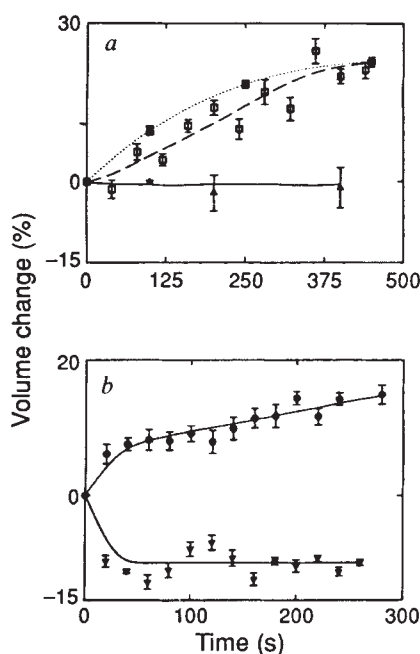


FIG. 2 Changes in protoplast volume induced by the photorelease of Ca^{2+} using caged probes, and by red light. Treatments in *a* were: 25 s ultraviolet photolysis of caged Ca^{2+} (■), 25 s ultraviolet photolysis of caged $InsP_3$ (□), and a control in which protoplasts were only exposed to green light of 488 nm from the argon laser (▲). Treatments in *b* were: 3 min red light (●) and 2 min red followed by 2 min far red light (▼). Protoplast volumes were determined from non-confocal bright field images of the same protoplasts used in Fig. 1, obtained with the transmission detector of the confocal microscope. The volumes of the spherical protoplasts were calculated using Semper 6 Plus image analysis software (Synoptics, Cambridge, UK). Protoplast immobilization was found to modify $[Ca^{2+}]_i$ and swelling. Protoplasts were therefore not adhered to the slide and consequently could not be perfused because this caused them to move. Incubation times in all cases were thus limited to 500 s, to ensure protoplasts maintained full viability. The zero time point has been taken as the end of illumination and commencement of $[Ca^{2+}]_i$ increase in each case where this occurred. Thus the total observation period for volume measurement for red, red/far red irradiated protoplasts and for photolysis of caged Ca^{2+} , and dark maintained protoplasts were 260, 240, 450 and 400 s, respectively. The number of replicates was as for Fig. 1 except for caged Ca^{2+} ($n=50$) and the control ($n=5$).

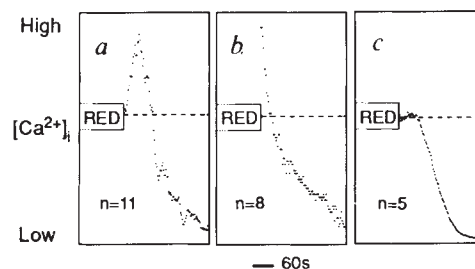


FIG. 3 Patterns of variation in the timing of red light-induced $[Ca^{2+}]_i$ changes in etiolated protoplasts. *a*, Protoplasts in which the $[Ca^{2+}]_i$ elevation began after the end of red light illumination. *b*, Protoplasts in which $[Ca^{2+}]_i$ elevation was initiated before the end of illumination and was above resting level when illumination ceased. *c*, Protoplasts which showed no apparent increase in $[Ca^{2+}]_i$. A single data point in the traces is an average pixel value from a confocal fluorescence image of a protoplast at each time point. A representative trace is shown for each of the three groups (*a-c*).

been drawn from Ca^{2+} signalling studies in animal cells^{18,19}. Because the prolonged Ca^{2+} efflux was not observed when Ca^{2+} was released from caged Ca^{2+} or caged $InsP_3$, this phase may represent some additional metabolic control specifically activated by red light. Ca^{2+} efflux is clearly not essential for induction of the physiological response of protoplast swelling although it may constrain the size of the response; red light resulted in a 15% increase in volume whereas photoactivation of caged Ca^{2+} or caged $InsP_3$ caused a 24% volume increase.

These results illustrate two important points. First, both the artificial elevation of $[Ca^{2+}]_i$ and of red light induce the physiological response of protoplast swelling. We therefore conclude that red light-mediated effects can be transduced through $[Ca^{2+}]_i$. These data concur with earlier observations⁶ which showed that red light-induced swelling could be inhibited by Ca^{2+} channel blockers and EGTA. Second, our data demonstrate the value of using caged probes for dissecting the molecular pathway between phytochrome and red light-induced responses. Releasing Ca^{2+} from its caged probe only influences one step of the signal transduction chain in a manner that is more pronounced than that induced by red light. The phytochrome-mediated red light signal also induces a significant efflux of Ca^{2+} , an effect which is apparently not relevant to the physiological response of protoplast swelling. At present, we are investigating whether the photo-release of Ca^{2+} from its caged probe inside protoplasts influences the expression of genes known to be red light-regulated³. Such observations could greatly clarify the details of the transduction pathway linking phytochrome with gene expression itself. □

Received 5 June; accepted 20 July 1992.

- Hart, J. W. *Light and Plant Growth* (eds Black, M. & Chapman, J.) (Unwin Hyman, London, 1988).
- Wagner, D., Tepperman, J. M. & Quail, P. H. *Pl. Cell* **3**, 1275–1288 (1991).
- Smith, H. in *Light and Plant Development* (ed. Smith, H.) 493–503 (Butterworths, London, 1976).
- Bossen, M. E., Dassen, H. H. A., Kendrick, R. E. & Vredenberg, W. J. *Planta* **174**, 94–100 (1988).
- Viner, N., Whitlam, C. & Smith, H. *Planta* **175**, 209–213 (1988).
- Beever, L., Loveys, B., Pearson, J. A. & Wareing, P. F. *Planta* **90**, 286–294 (1970).
- Zhou, S., Jones, A. M. & Scott, T. M. *Pl. Cell Rep.* **9**, 435–438 (1990).
- Minta, A., Kao, J. P. Y. & Tsien, R. Y. *J. Biol. Chem.* **264**, 8171–8178 (1989).
- Kao, J. P. Y., Harootunian, A. T. & Tsien, R. Y. *J. Biol. Chem.* **264**, 8179–8184 (1989).
- Gilroy, S., Read, N. D. & Trewavas, A. J. *Nature* **346**, 769–771 (1990).
- Read, N. D. *et al.* *J. Microsc.* **165**, 56–66 (1992).
- Thiel, G., Macrobbe, E. A. C. & Hanke, D. *EMBO J.* **9**, 1737–1741 (1990).
- Grynciewicz, G., Pooenie, R. Y. & Tsien, R. Y. *J. Biol. Chem.* **260**, 3440–3450 (1989).
- Gilroy, S., Fricker, M. D., Read, N. D. & Trewavas, A. J. *Pl. Cell* **3**, 333–344 (1991).
- Sage, S. O., Merritt, J. E., Hallam, T. J. & Rink, T. J. *Biochem. J.* **258**, 923–926 (1989).
- Cheek, T. R., Burgoyne, R. D. & Berridge, M. J. in *Cell Signalling: Experimental Strategies* (eds Reid, E., Cook, G. M. W. & Luzzio, J. P.) 287–302 (Royal Society of Chemistry, Cambridge, 1991).
- Hale, C. C. & Roux, S. J. *Pl. Physiol.* **65**, 658–662 (1980).
- Millard, P. J., Gross, D., Webb, M. W. & Fawcett, C. *Proc. natn. Acad. Sci. U.S.A.* **85**, 1854–1858 (1988).
- O'Sullivan, A. J., Cheek, T. J., Moreton, R. B., Berridge, M. J. & Burgoyne, R. D. *EMBO J.* **8**, 401–411 (1989).

ACKNOWLEDGEMENTS. We thank the Science and Engineering Research Council for grant funding and a studentship to P.S.S.

Mixed parvocellular and magnocellular geniculate signals in visual area V4

Vincent P. Ferrera, Tara A. Nealey
& John H. R. Maunsell*

Department of Physiology and Center for Visual Science,
University of Rochester, Rochester, New York 14642-8642, USA

VISUAL information from the retina is transmitted to the cerebral cortex by way of the lateral geniculate nucleus (LGN) in the thalamus. In primates, most of the retinal ganglion cells that project to the LGN belong to one of two classes, P and M, whose axons terminate in the parvocellular or magnocellular subdivisions of the LGN. These cell classes give rise to two channels that have been distinguished anatomically, physiologically and behaviourally^{1-3,16,17}. The visual cortex also can be subdivided into two pathways, one specialized for motion processing and the other for colour and form information⁴. Several lines of indirect evidence have suggested a close correspondence between the subcortical and cortical pathways, such that the M channel provides input to the motion pathway and the P channel drives the colour/form pathway⁵⁻⁷. This hypothesis was tested directly by selectively inactivating either the magnocellular or parvocellular subdivision of the LGN and recording the effects on visual responses in the cortex. We have previously reported that, in accordance with the hypothesis, responses in the motion pathway in the cortex depend primarily on magnocellular LGN⁸. We now report that in the colour/form pathway, visual responses depend on both P and M input. These results argue against a simple correspondence between the subcortical and cortical pathways.

We examined P and M channel contributions to cortical responses in the colour/form pathway by measuring their influence on the responses of neurons in visual area V4. We did this by taking advantage of the anatomical segregation of the P and M channels into different layers in the LGN. We selectively inactivated either one channel or the other by injecting a small amount of a pharmacological agent (either lidocaine or GABA (γ -aminobutyric acid)) that temporarily abolishes neuronal responses, into a single layer in the LGN⁹. By comparing the strength of cortical responses to a visual stimulus before and during a selective LGN block, we could determine whether the cortical response depended on a contribution from the blocked channel.

We recorded the visual responses of single units or small clusters of units in area V4 of four anaesthetized macaque monkeys (two each of *Macaca fascicularis* and *M. nemestrina*). Figure 1 (left panels) shows the response of a multiunit site in V4 before and after selectively blocking the magnocellular subdivision of the LGN. This block abolished the cortical response, indicating that this site depended on the M pathway for its visual response. Magnocellular blockade eliminated responses at about 10% of V4 sites. We quantified the effects of LGN injections with a blocking index that measures the proportion of the cortical response that is eliminated by an LGN injection. On average, magnocellular injections reduced responses in V4 by about 40% (average blocking index 0.38 ± 0.08 s.e.m.; $n = 44$).

As expected, parvocellular injections also reduced V4 responses (Fig. 1, right panels). Complete blocks were never observed, but this may reflect a failure to eliminate P channel contributions completely. The LGN is a six-layered structure with two magnocellular layers and four parvocellular layers. The contralateral visual stimuli activated only one magnocellular layer (layer 1), but two parvocellular layers (layers 4 and 6). Only one of the parvocellular layers could be blocked with an

injection. Blocking one or the other of the active parvocellular layers reduced responses in V4 ~30% (average blocking index 0.28 ± 0.06 s.e.m.; $n = 61$). At four sites we blocked both layers 4 and 6 simultaneously by using two injection probes. The mean reduction for paired parvocellular blocks was no greater than that for single injections (mean blocking index 0.24 ; $n = 4$), and none of these paired parvocellular blocks eliminated the visual response in V4.

These results show that both subcortical channels contribute to responses in V4, but questions remain about whether these contributions are segregated within V4. If P and M channel inputs were segregated in this area, one might expect to find sites that receive input almost exclusively from one channel, the inactivation of which would produce a complete block like that shown in Fig. 1 (left). This result was rare. Twenty-eight sites in V4 were tested first for contributions from one subcortical channel and then from the other. Only two sites had nearly complete blocks when one channel was inactivated (one M and one P) and had no obvious contribution from the other channel. It was common for a block of either channel to have an incom-

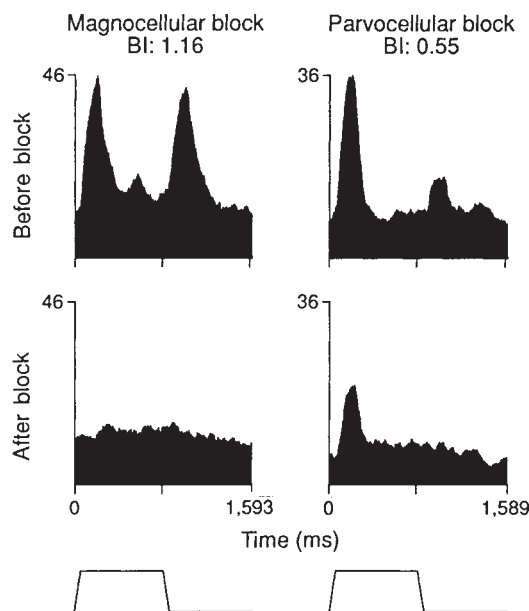


FIG. 1 Effects of magnocellular (left) and parvocellular (right) blockade on responses of two different multiunit sites in V4. The stimulus was an oscillating white bar on a dark background. The size and excursion of the bar were limited so that the stimulus stayed well within the visual field representation that was blocked in the LGN (see ref. 8). Other properties of the stimulus were matched to the selectivity of the recording site. Visual stimulation was provided to the contralateral eye so that only LGN layers 1, 4 and 6 were active. The line at the bottom represents stimulus position as a function of time for one stimulus cycle (the nearly vertical segments indicate movement to or from, the horizontal segments indicate stationary pauses). The vertical axis is calibrated in impulses per second. The impulse histograms were averaged over 50 stimulus cycles and smoothed by boxcar averaging. The top panels show the response averaged over the 50 stimulus cycles immediately before the injection of a 2% lidocaine solution into layer 1 (left of figure) or layer 6 (right) of the LGN. The bottom panels show the response averaged over the 50 stimulus cycles following the end of the injection period. We also collected 50 cycles of background activity without any stimulus. We calculated a response by taking the difference between the value of the largest and smallest bins in the averaged histograms. We then calculated a blocking index (BI) defined as $1.0 - (\text{post-injection response} - \text{background}) / (\text{pre-injection response} - \text{background})$. The blocking index for the magnocellular injection was 1.16, indicating that the response after the injection was completely abolished, and that fluctuations in activity were slightly less than those during the background period. The blocking index for the parvocellular injection was 0.55, yet this was among the clearest effects we obtained for injections that blocked a single parvocellular layer.

* To whom correspondence should be addressed.

plete effect, even though the amount of blocking agent injected was sufficient to abolish all neuronal activity at the injection site for several minutes. Figure 2 shows the responses of a multiunit site in V4 recorded while individually blocking each of the three active LGN layers. Each layer made a partial contribution to the response at this site.

Although M and P inputs appear to be mixed within small regions, they nevertheless might be separated on a cell by cell basis, with individual neurons receiving exclusively one input or the other. The limited single unit data we have collected so far ($n=7$) argues against this idea. Only one unit had a near-complete block following a parvocellular injection, and none were unresponsive following magnocellular injections. Most units had only a partial reduction in response when either channel was inactivated. Thus, whereas some units in V4 may receive exclusively magnocellular or parvocellular input, the majority appear to receive mixed inputs.

The failure of either subcortical channel to dominate responses in the colour/form pathway stands in contrast to the situation in the motion pathway. In a previous study we found that responses in the middle temporal visual area (MT), in the motion pathway, depended primarily on magnocellular input, although a few MT sites showed significant effects of parvocellular blockade⁸. It appears that although MT receives input predominantly from one subcortical channel, V4 does not, although both lie at the same level in the hierarchy of cortical areas defined by anatomical connections¹⁰. This difference between the two areas is illustrated in Fig. 3, which compares the average blocking indices for V4 with those obtained previously for MT. No parallel to the striking imbalance between magnocellular and parvocellular contributions to MT was found in area V4. Both sets of experiments were conducted under nearly identical conditions. Additional experiments were done to ensure that the effects of magnocellular LGN injections were

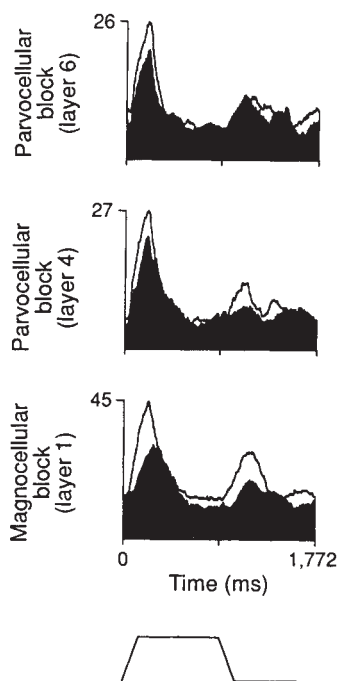


FIG. 2 Responses from one V4 multiunit site recorded before and after three different injections that individually blocked each of the active layers of the LGN. As in Fig. 1, responses were averaged over 50 stimulus cycles before each injection (outlined histograms) and 50 cycles immediately following the injection (filled histograms). The vertical axis represents impulses per second. Each LGN layer contributed a portion of the cortical response. The blocking indices were 0.40 for layer 6, 0.29 for layer 4 and 0.50 for layer 1.

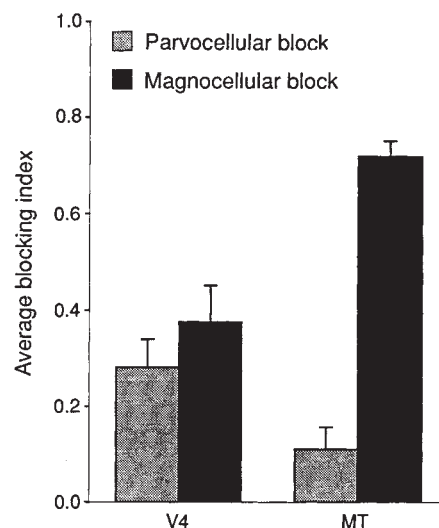


FIG. 3 Average blocking indices for area V4 and MT. We tested 61 V4 sites with parvocellular injections (average blocking index 0.28 ± 0.06 s.e.m.) and 44 sites with magnocellular injections (average blocking index 0.38 ± 0.08 s.e.m.). For MT, we tested 75 sites with parvocellular injections (average blocking index 0.11 ± 0.05 s.e.m.) and 57 sites with magnocellular injections (average blocking index 0.72 ± 0.03 s.e.m.). These averages include single and multiunit cortical sites and different blocking agents. For V4, the LGN blocks were made with lidocaine (2%) or GABA (25 mM). There was no significant difference between lidocaine and GABA for either parvocellular or magnocellular injections. The MT data were collected using lidocaine and $MgCl_2$. It is unlikely that the effects of magnocellular injections were due to spread of the blocking agent into the parvocellular layers. For this to occur, an injection into layer 1 would have to spread through layers 2 and 3, which were inactive. In control experiments we measured the spread of the injection by recording at varying distances from the injection site. There was no effect of any blocking agent on LGN responses recorded 1.5 mm from the injection site. We have also recorded from the parvocellular neurons representing the stimulus while blocking the magnocellular layers (or vice versa) and saw no spread of effects within the LGN (see ref. 8).

not due to spread of the blocking agent into the parvocellular layers (see Fig. 3 legend).

Because the M and P channels have different stimulus selectivities, their relative contributions are likely to depend on the spatial and temporal characteristics of the stimulus as well as its contrast and colour. We used small drifting or flashing bars that had sharp edges and high luminance contrast. These stimuli produced strong responses from both magnocellular and parvocellular neurons in the LGN, and we do not believe that they systematically favoured either the P or M channel. It is possible that the relative contributions to cortical responses would appear different using stimuli that decidedly favoured one or the other subcortical channel.

The routes by which M channel signals reach V4 are not yet known. M channel input could be mediated by the reciprocal connections between V4 and MT¹⁰. It is also possible that the inputs to V4 reflect mixing of the M and P channels in the superficial layers of V1. Malpeli *et al.*¹¹ used selective LGN inactivation to demonstrate that many single units in V1 receive both M and P channel input. We have shown that the cytochrome oxidase blobs and interblobs of V1, which provide inputs to V4, both receive substantial input from the M channel¹².

Our results provide direct evidence that both M and P channels make substantial contributions to neuronal responses in the colour/form pathway. Furthermore, we found no evidence that M and P channel contributions are segregated within area V4. Thus, there does not seem to be a one-to-one correspondence between the M and P channels and the cortical pathways. The hypothesized correspondence between the cortical and subcortical pathways⁵⁻⁷ was based on indirect evidence which is

consistent with our observations. Indeed, intermixing of M and P channel contributions in cortex had been suggested by a variety of earlier observations^{10,11,13-18}.

Although the correspondence between the cortical and subcortical pathways is not one-to-one, the dominance of the motion pathway by the M channel (if neurons in MT are representative of that pathway) suggests that distinct relationships exist between the subcortical and cortical pathways. Overall, the available data suggest an asymmetric relationship between cortical and subcortical pathways in which the cortical motion pathway is largely dominated by the M channel, whereas the colour/form pathway receives inputs from both the P and M channels. □

Received 30 March; accepted 30 June 1992.

- Shapley, R. & Perry, V. H. *Trends Neurosci.* **9**, 229-235 (1986).
- Silveira, L. C. L. & Perry, V. H. *Neuroscience* **40**, 217-237 (1991).
- Schiller, P. H., Logothetis, N. K. & Charles, E. R. *Vis. Neurosci.* **5**, 321-346 (1990).
- Ungerleider, L. G. & Mishkin, M. *Analysis of Visual Behavior* (eds Ingle, D. J., Goodale, M. A. & Mansfield, R. J. W.) (MIT Press, Cambridge, MA, 1982).
- Livingstone, M. S. & Hubel, D. H. *Science* **240**, 740-749 (1988).
- Maunsell, J. H. R. *Matters of Intelligence* (ed. Vaina, L. M.) (Reidel, Dordrecht, 1987).
- Maunsell, J. H. R. & Newsome, W. T. A. *Rev. Neurosci.* **10**, 363-401 (1987).
- Maunsell, J. H. R., Nealey, T. A. & DePriest, D. D. *J. Neurosci.* **10**, 3323-3334 (1991).
- Malpeli, J. G. & Schiller, P. H. *J. Neurosci. Meth.* **1**, 143-157 (1979).
- Felleman, D. J. & Van Essen, D. C. *Cerebral Cortex* **1**, 1-47 (1991).
- Malpeli, J. G., Schiller, P. H. & Colby, C. L. *J. Neurophysiol.* **46**, 1102-1119 (1981).
- Nealey, T. A. & Maunsell, J. H. R. *Invest. Ophthalmol. Vis. Sci. Suppl.* **32**, 1117 (1991).
- Blasdel, G. G., Lund, J. S. & Fitzpatrick, D. *J. Neurosci.* **5**, 3350-3369 (1985).
- Fitzpatrick, D., Lund, J. S. & Blasdel, G. G. *J. Neurosci.* **5**, 3329-3349 (1985).
- Lachica, E. A., Beck, P. D. & Casagrande, V. A. *Proc. natn. Acad. Sci. U.S.A.* **89**, 3566-3570 (1992).
- Merigan, W. H., Byrne, C. & Maunsell, J. H. R. *J. Neurosci.* **11**, 3422-3429 (1991).
- Merigan, W. H., Katz, L. M. & Maunsell, J. H. R. *J. Neurosci.* **11**, 994-1001 (1991).
- Schiller, P. H., Logothetis, N. K. & Charles, E. R. *Vis. Neurosci.* **5**, 321-346 (1990).

ACKNOWLEDGEMENTS. We thank R. A. Eatock, W. H. Merigan and P. Lennie for comments on earlier versions of this manuscript and P. G. Vamvakias for technical assistance. The research was supported by an award from the National Eye Institute. V.P.F. was supported by a fellowship from the National Institutes of Health.

Uncoupling of hypomyelination and glial cell death by a mutation in the proteolipid protein gene

Armin Schneider*, Paul Montague*†, Ian Griffiths†, Monica Fanarraga†, Peter Kennedy†, Peter Brophy‡ & Klaus-Armin Nave*§

* Zentrum für Molekulare Biologie (ZMBH), Universität Heidelberg, D-6900 Heidelberg, Germany

† Applied Neurobiology Group, University of Glasgow, Bearsden, Glasgow G61 1QH, UK

‡ Department of Biological and Molecular Sciences, University of Stirling, Stirling FK9 4LA, UK

PROTEOLIPID protein (PLP; M_r 30,000) is a highly conserved major polytopic membrane protein in myelin but its cellular function remains obscure. Neurological mutant mice can often provide model systems for human genetic disorders. Mutations of the X-chromosome-linked PLP gene are lethal, identified first in the jimpy mouse^{1,2} and subsequently in patients with Pelizaeus-Merzbacher disease^{3,4}. The unexplained phenotype of these mutations includes degeneration and premature cell death of oligodendrocytes with associated hypomyelination⁵. Here we show that a new mouse mutant *rumpshaker*⁶ is defined by the amino-acid substitution Ile-to-Thr at residue 186 in a membrane-embedded domain of PLP. Surprisingly, *rumpshaker* mice, although myelin-deficient, have normal longevity and a full complement of morphologically normal oligodendrocytes⁷. Hypomyelination can thus be genetically separated from the PLP-dependent oligodendrocyte degeneration. We suggest that PLP has a vital function in glial

cell development, distinct from its later role in myelin assembly, and that this dichotomy of action may explain the clinical spectrum⁸ of Pelizaeus-Merzbacher disease.

Hypomyelination in the *rumpshaker* mouse⁶ (*rsh*) is confined to the central nervous system (CNS) and the brain is affected more severely than the spinal cord. The *rsh* optic nerve contains numerous unmyelinated axons with intervening astrocytic processes (Fig. 1a), showing similarities to the *jimpy* (*jp*) mouse⁹. But unlike *jp*, hypomyelination is not caused by a lack of oligodendrocytes^{9,10}. Morphometric analysis⁷ of the optic nerve and spinal cord shows that their number is, in fact, slightly increased and there are no signs of degeneration. Electron microscopy shows a prominent rough endoplasmic reticulum and Golgi apparatus⁷, suggesting that the oligodendrocytes are metabolically active. Additionally, myelin sheaths from *rsh* mutants immunostain strongly and selectively for the DM-20 isoform of PLP (Fig. 1c, d) (DM-20, M_r 26,000, is translated from an alternatively spliced PLP transcript¹¹ that appears much earlier in development¹²).

We investigated the relationship between the X-chromosome-linked *rsh* gene and the PLP gene in more detail. PLP messenger RNAs appeared unchanged in size (Fig. 2a) and significantly more abundant than PLP mRNA in age-matched *jp* mutant mice (in which a mutation of the PLP gene^{1,2} causes oligodendrocyte death). Using RNase protection analysis, we found an mRNA ratio of 3.3 (PLP/DM 20) in adult normal mice, but only 1.1 in mutant mice (Fig. 2b). This reduction of PLP mRNA has been confirmed by a 50% decrease in immunostained PLP, and an increase in relative abundance of the DM-20 isoform by three times, in 20-day-old *rsh* mice, as monitored by western blot analysis of myelin-enriched membrane fractions¹³.

We suspected that the mutation interferes primarily with oligodendrocyte differentiation (possibly by abnormal PLP mRNA splicing) and that the virtual absence of PLP from compact myelin is, in part, a post-translational phenomenon. Using primers against the 5' and 3' untranslated regions, we cloned both PLP and DM-20 from brain complementary DNA

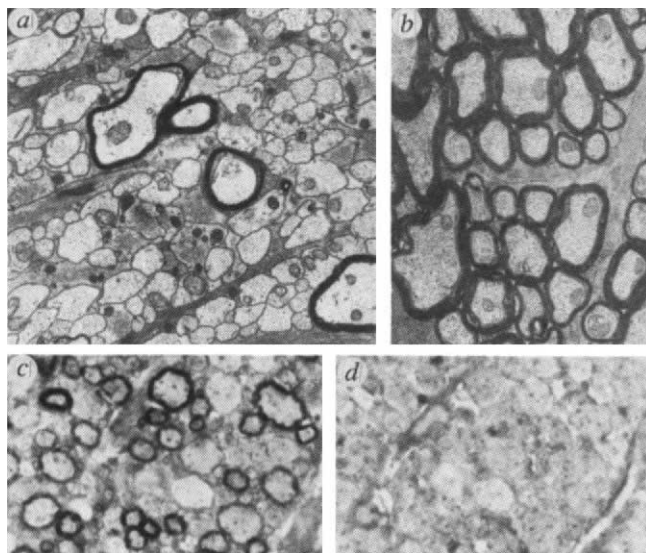


FIG. 1 Optic nerves from a, 90-day-old mutant and b, wild-type mice. Most *rsh* axons are unmyelinated; the occasional sheaths present are disproportionately thin. Astrocyte processes are increased (magnification $\times 7,500$). c, d, Adjacent 1 μ m resin sections from the *rsh* spinal cord immunostained by the peroxidase-antiperoxidase method using antiserum against a C-terminal peptide (residues 271-276) common to both PLP and DM-20 (c) or with a PLP-specific (residues 117-129) antiserum (d). Many axons lack a myelin sheath but the majority of sheaths present are immunostained by the C-terminal anti-peptide antiserum, whereas the identical sheaths are unstained by the PLP-specific antiserum.

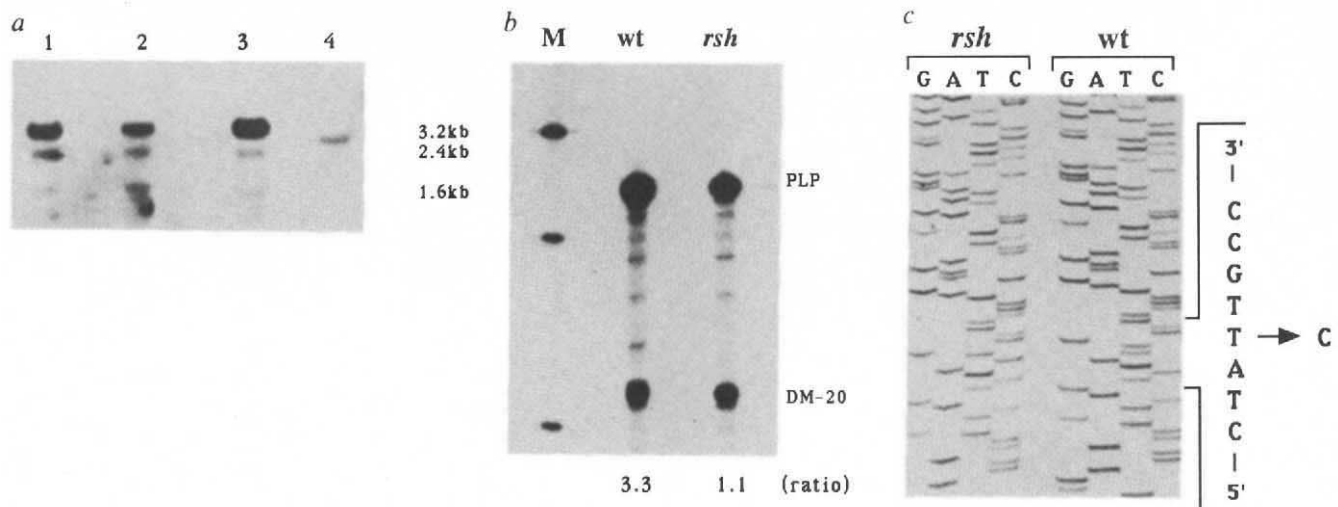


FIG. 2 Expression of mutant PLP transcripts in *rsh* mice. **a**, Northern blot analysis of PLP mRNA from normal male C3H $\times$ 101/H mice (lane 1), hemizygous *rsh*/Y (C3H $\times$ 101/H) mice (lane 2), normal male (B6 $\times$ CBA) mice (lane 3), and hemizygous *jp*/Y (B6 $\times$ CBA) mice (lane 4). There are three alternatively polyadenylated PLP mRNAs of the indicated size²⁴. **b**, Alternative PLP mRNA splicing was monitored by RNase protection analysis, generating fragments specific for PLP (393 bp) and DM-20 (238 bp) in normal littermates (lane wt) and *rsh* mutants (lane *rsh*). The calculated mRNA ratio (PLP/DM-20) is shown below. **c**, Sequence analysis of cloned *rsh* and normal PLP cDNA. Shown to the right is the normal sequence with the single nucleotide difference identified in codon 186 of exon IV.

METHODS. All *rsh* mice on a C3H $\times$ 101/H background were from our own colony, and *jp* breeding pairs (B6 $\times$ CBA) obtained from the Jackson Laboratories. For northern analysis, 10 μ g of total brain RNA (postnatal day 25) was size fractionated on 1% agarose gels, transferred to Nytran membranes (Schleicher and Schüll) and probed with a random prime-labelled

mouse PLP cDNA fragment². Equal RNA loading was verified by methylene blue staining (not shown). For RNase protection, the 393 bp *Bgl*III-*Pvu*II fragment² of pC4 was subcloned into the *Eco*RV/*Bam*HI site of pSP72 (Promega) and linearized with *Xho*I. [³²P]-UTP-labelled antisense RNA was hybridized to 50 μ g total brain and liver RNA as described¹¹. Autoradiograms were densitometrically scanned (Hirschmann EL400) and readings were normalized for incorporated ³²P-UTP. For sequence analysis, PLP and DM-20 cDNAs were amplified by PCR in triplicate from separate batches of random-primed brain cDNA, using primers against the 5' untranslated region (5'-CAGGATCCTCCAGCTGAGCAAAGTC-3') and 3' untranslated region (5'-ATGGATCCACAAAGGGAGTTCTATGGAGCTC-3') of PLP mRNA²⁴. PCR products were gel-purified, digested with *Bam*HI, and cloned into pBluescript (Stratagene). Sequence analysis was carried out with internal primers using a T7 dideoxy-sequencing kit (Pharmacia) according to the manufacturer's protocol. Each sequence was determined from three unrelated clones.

of *rsh* mutants and normal littermates. For sequence analysis, each cDNA was amplified by the polymerase chain reaction and cloned into a plasmid. We also amplified and subcloned the PLP promoter encompassing about 700 base pairs (bp) of 5' regulatory region from normal and mutant genomic DNA. As shown in Fig. 2c, only a single nucleotide difference was identified: codon 186 in exon IV was changed from ATT (Ile) in the wild-type cDNAs to ACT (Thr) in the *rsh* cDNAs. All other sequences from *rsh* mice were identical to those from normal littermates, including the alternative 5' splice site within exon III, and encoded an otherwise normal^{11,14} mouse PLP and DM-20.

The T $\rightarrow$ C transition in exon IV generated an *Acc*I restriction site (Fig. 3a) which was confirmed by genomic Southern analysis. Using a cDNA probe for the PLP gene, we detected no major deletion or rearrangement of the *rsh* PLP gene. But an *Acc*I digest resulted in two novel, *rsh*-specific 4.4-kilobase (kb) fragments (a doublet in Fig. 3b). To rule out the possibility that these fragments represent a mouse restriction-fragment-length polymorphism (RFLP), we also established the genotype of the original parental inbred strains, C3H and 101/H (Fig. 3c). Exon IV was amplified from genomic DNA using primers flanking codon 186. The parental strains did not contain the internal *Acc*I site, suggesting that we have identified the *rsh* mutation in the PLP gene.

The substitution Ile $\rightarrow$ Thr affects the non-polar M4 domain of PLP assumed to be a membrane-embedded α -helix (Fig. 4; for a discussion of alternative topological models, see ref. 15). Several previously identified PLP mutations are also single-residue changes in transmembrane domains, but not in M4 (ref. 4). These alleles are lethal disorders characterized by the degeneration and death of oligodendrocytes as the primary cause of hypomyelination, as documented best for the *jp* mouse^{10,16-19}, which is defined by the complete lack² of M5. The allelic

mutation *jp*^{msd} is defined by the amino-acid substitution Ala $\rightarrow$ Val at residue 242 (ref. 20) in M5 (Fig. 4) yet displays a very similar phenotype to *jp*, including the lack of mature oligodendrocytes²¹. In contrast to *jp*, *jp*^{msd} or any other mutant of the PLP gene, the *rsh* mouse demonstrates that changes of the PLP structure can cause marked hypomyelination unassociated with cell death. Moreover, the inability to assemble myelin cannot be an underlying cause of oligodendrocyte death, as *rsh* mice are clearly myelin-deficient. We do not understand the functional difference between PLP^{Thr186} and other substitutions in the PLP gene causing this unique non-lethal phenotype. It seems most likely that substitutions such as PLP^{Val242} (in *jp*^{msd}) or PLP^{Thr186} (in *rsh*) interfere with different interactions of the membrane-embedded PLP domains, possibly in the formation of a functional oligomeric structure.

Mutations of the PLP gene demonstrate important differences to previously described genetic defects causing abnormal cell death. For example in *Caenorhabditis elegans*, some mutations cause degenerative changes distinct from programmed cell death²², such as the dominant mutations in the *mec-4* gene expressed in touch-receptor neurons²³ (presumably because a mutant membrane protein can be cytotoxic). In this case, neuronal degeneration occurs late in development, and the recessive null mutation and the majority of the *mec-4* alleles are not degenerative. In contrast, all PLP mutations except *rsh* are lethal, and oligodendrocytes show very early signs of degeneration even before PLP would become immunodetectable¹⁹. Comparing oligodendrocytes in *rsh* mice with those from *jp* or *jp*^{msd} mice, we propose that PLP has a critical function in these cells before myelin assembly. This function is lost in most PLP mutants but sufficiently preserved in PLP^{Thr186}, enabling *rsh* oligodendrocytes to survive and differentiate. The early expression of DM-20, and its unexplained relative abundance in *rsh*, suggest it as a candidate for this critical role, however

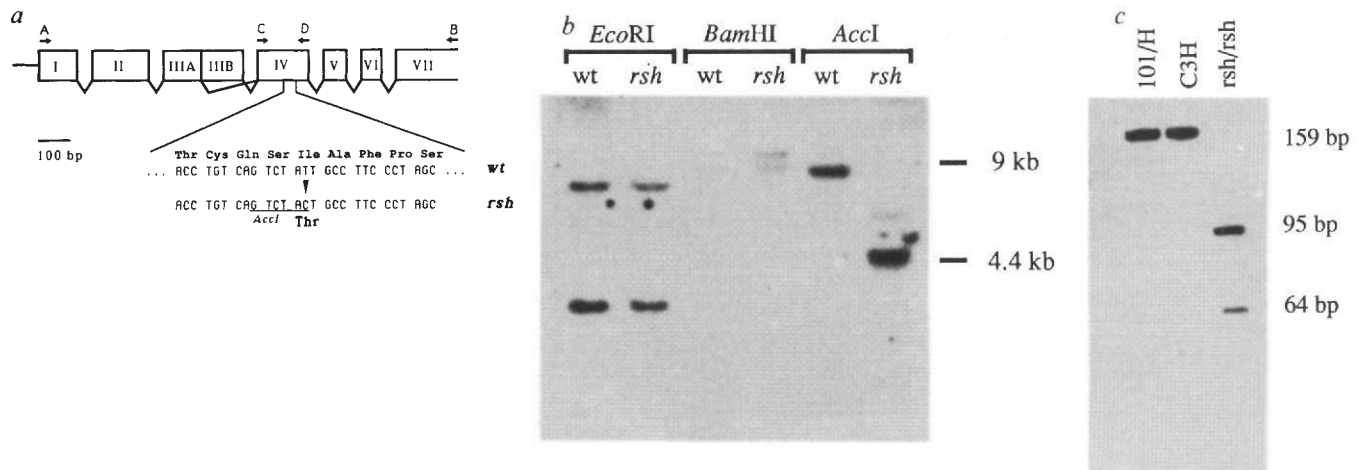


FIG. 3 *a*, Organization of the mouse PLP gene with alternative splice sites in exon III and the *rsh* mutation in exon IV. Part of the open reading frame of the normal gene and the *rsh* mutation (below) that generates a new restriction site is shown in detail. The location of primer pairs used for cDNA cloning (A and B) and amplification of exon IV (C and D) is indicated by arrows. Intron sizes not to scale. *b*, Southern analysis of genomic DNA from mutant mice (*rsh*) and normal littermates (wt) digested with the indicated restriction enzymes and probed with a PLP cDNA. The shorter AccI fragment of the *rsh* PLP gene represents a doublet of about 4.4 kb in size. *c*, Restriction analysis of exon IV from *rsh* mice and both parental inbred strains (C3H and 101/H). The undigested PCR product is 159 bp long, yielding

two *rsh* specific AccI fragments of 95 bp and 64 bp.

METHODS. Southern analysis was carried out with 5 µg liver genomic DNA by standard protocols³¹. Duralon membranes (Stratagene) were probed with a random prime-labelled mouse PLP cDNA (*Bam*HI-*Nco*I of pC4)² and washed with 0.2 × SSC at 65 °C. For restriction analysis, exon IV was amplified by PCR with primers C (5'-CATCACCTATGCCCTGA-3') and D (5'-TACATTCTGG-CATCAGCGCAGAGACTGC-3') using tail genomic DNA from homozygous *rsh* mutants and mice of the original parental inbred strains (provided by the MRC Radiobiology Unit). PCR products were trace-labelled with ³²P-dCTP and digested with AccI before separation on a 5% acrylamide gel and autoradiography.

the nature of this 'premyelin' function is unknown.

The unusual high conservation of the PLP protein (100% between mouse, rat²⁴ and human²⁵) suggests multiple protein-protein interactions and possibly the functional oligomerization of PLP. When reconstituted into a lipid bilayer, the purified protein mediates the transmembrane flux of ²²Na⁺ ions²⁶. It is tempting to speculate that PLP is a receptor or channel protein of developing oligodendroglial cells which has only been recruited as a structural membrane component later in vertebrate evolution. Lack of the 'premyelin' function of PLP may be partly overcome by exogenous factors, since in the presence of conditioned medium cultivated *jimpy* oligodendrocytes are stimulated to differentiate far beyond their normal potential²⁷. The dichotomy of PLP action may be similar to the function of a

structurally related myelin protein of the peripheral nervous system, PMP-22/gas3, first identified as a growth-arrest regulated gene²⁸ and recently implicated in the Schwann cell defect and dysmyelination of *Trembler* mutant mice²⁹ and in patients with Charcot-Marie-Tooth disease type 1a (ref. 30).

Human Pelizaeus-Merzbacher disease is a fatal disorder and has been grouped into several subclasses⁸ based on a considerable clinical range of the disease. We propose that each point mutation of the PLP gene results in a loss of function that is position-dependent and affects either the 'premyelin' or the structural function of PLP or both. It is possible that these X-chromosome-linked diseases are dominant at the cellular level as 'normal' heterozygotes are mosaics. This question can now be experimentally addressed by transgenic expression of the wild-type PLP gene in *jimpy* and *rumpshaker* mice. □

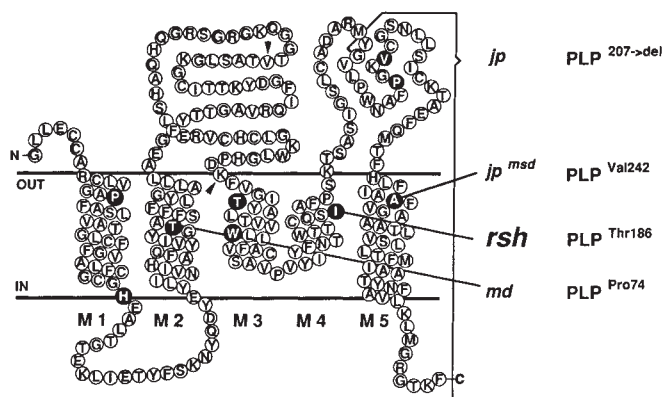


FIG. 4 Hypothetical model of PLP as a polytopic membrane protein showing the effect of the *rsh* mutation (PLP^{Thr186}). Also shown are the lethal mutations in the *jimpy* mouse (a deletion of the carboxyl terminus²), in the *jp^{msd}* allele²⁰, and in the *md* rat³². The positions of other single-residue substitutions including several families with Pelizaeus-Merzbacher disease⁴, are indicated by black circles. The wild-type sequence (identical in mouse, rat and human) is shown. Putative α-helical membrane-associated domains of PLP are numbered M1 to M5. The position of residues 116-150, absent in the DM-20 isoform, is indicated by two arrowheads.

Received 14 April; accepted 9 July 1992.

1. Dautigny, A. et al. *Nature* **321**, 867-869 (1986).
2. Nave, K.-A., Lai, C., Bloom, F. E. & Milner, R. J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9264-9268 (1986).
3. Hudson, L. D., Puckett, C., Berndt, J., Chan, J. & Gencic, S. *Proc. natn. Acad. Sci. U.S.A.* **86**, 8128-8131 (1989).
4. Hudson, L. D. *Myelination and Dysmyelination* (eds Duncan, I. D., Skoff, R. P. & Colman, D.) 155-165 (New York Academy of Sciences, New York, 1990).
5. Nave, K.-A. & Milner, R. J. *CRC crit. Rev. Neurobiol.* **5**, 65-91 (1989).
6. Griffiths, I. R. et al. *J. Neurocytol.* **19**, 273-283 (1990).
7. Fanarraga, M. L. et al. *Glia* **5**, 161-170 (1992).
8. Seitelberger, F. *Handbook of Clinical Neurology* (eds Klinken, P. J. & Bruyn, G. W.) 150-202 (North-Holland, Amsterdam, 1970).
9. Skoff, R. P. *Nature* **264**, 560-562 (1974).
10. Knapp, P. E., Skoff, R. P. & Redstone, D. W. *J. Neuroscience* **6**, 2813-2822 (1986).
11. Nave, K.-A., Lai, C., Bloom, F. E. & Milner, R. J. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5665-5669 (1987).
12. Ikenaka, K., Kagawa, T. & Mikoshiba, K. *J. Neurochem.* **58**, 2248-2253 (1992).
13. Mitchell, L. S. et al. *J. Neurosci. Res.* (in press).
14. Hudson, L. D., Berndt, J. A., Puckett, C., Cozak, C. A. & Lazzarini, R. A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1454-1458 (1987).
15. Popot, J.-L., Pham-Dinh, D. & Dautigny, A. *J. Membrane Biol.* **120**, 233-246 (1991).
16. Sidman, R. L., Dickie, M. M. & Appel, S. H. *Science* **144**, 309-311 (1964).
17. Meyer, C. & Bischoff, A. *J. Neuropathol. exp. Neurol.* **33**, 343-353 (1974).
18. Billings-Gagliardi, S., Adcock, L. H. & Wolf, M. K. *Brain Res.* **194**, 325-338 (1980).
19. Vermeesch, M. K., Knapp, P. E., Skoff, R. P., Studzinski, D. M. & Benjamins, J. A. *Dev. Neurosci.* **12**, 303-315 (1990).
20. Gencic, S. & Hudson, L. D. *J. Neurosci.* **10**, 117-124 (1990).
21. Wolf, M. K., Kardon, G. B., Adcock, L. H. & Billings-Gagliardi, S. *Brain Res.* **271**, 121-129 (1983).
22. Raff, M. C. *Nature* **356**, 397-400 (1992).
23. Chalfie, M. & Wolinsky, E. *Nature* **345**, 410-416 (1990).
24. Milner, R. J. et al. *Cell* **42**, 931-939 (1985).
25. Diehl, H.-J., Schaich, M., Budzinski, R.-M. & Stoffel, W. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9807-9811 (1986).

26. Helynick, G. *et al.* *Eur. J. Biochem.* **133**, 689–695 (1983).
27. Bartlett, W. P., Knapp, P. & Skoff, R. P. *Glia* **1**, 253–259 (1988).
28. Manfioletti, G., Ruaro, M. E., Del Sal, G., Phillipson, L. & Schneider, C. A. *Molec. cell. Biol.* **10**, 2924–2930 (1990).
29. Suter, U. *et al.* *Nature* **356**, 241–244 (1992).
30. Timmerman, V. *et al.* *Nature Genet.* (in the press).
31. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning* (Cold Spring Harbor Laboratory, New York, 1989).
32. Boison, D. & Stoffel, W. *EMBO J.* **8**, 3295–3302 (1989).

ACKNOWLEDGEMENTS. We thank H. Krischke for technical assistance, and C.-M. Becker and members of our laboratory for discussion and critical reading of the manuscript. We also thank J. Nussbaum (Strasbourg) and N. Groome (Oxford) for providing anti-peptide antibodies. The study was supported by the Multiple Sclerosis Society of Scotland and the Bundesministerium für Forschung und Technologie (BMFT).

Processing of mutant cystic fibrosis transmembrane conductance regulator is temperature-sensitive

Gerene M. Denning*, Matthew P. Anderson*, Jane F. Amara†, John Marshall†, Alan E. Smith† & Michael J. Welsh*

* Howard Hughes Medical Institute, University of Iowa College of Medicine, Departments of Internal Medicine and Physiology and Biophysics, Iowa City, Iowa 52242, USA

† Genzyme Corporation, Framingham, Massachusetts 01701, USA

CYSTIC fibrosis transmembrane conductance regulator (CFTR) is a plasma membrane Cl^- channel regulated by cyclic AMP-dependent phosphorylation and by intracellular ATP^{1-7} . Mutations in CFTR cause cystic fibrosis⁸⁻¹⁰ partly through loss of cAMP-regulated Cl^- permeability from the plasma membrane of affected epithelia^{11,12}. The most common mutation in cystic fibrosis is deletion of phenylalanine at residue 508 (CFTR Δ F508) (ref. 10). Studies on the biosynthesis^{13,14} and localization¹⁵ of CFTR Δ F508 indicate that the mutant protein is not processed correctly and, as a result, is not delivered to the plasma membrane. These conclusions are consistent with earlier functional studies which failed to detect cAMP-stimulated Cl^- channels in cells expressing CFTR Δ F508 (refs 16, 17). Chloride channel activity was detected, however, when CFTR Δ F508 was expressed in *Xenopus* oocytes¹⁸, Vero cells¹⁹ and Sf9 insect cells²⁰. Because oocytes and Sf9 cells are typically maintained at lower temperatures than mammalian cells, and because processing of nascent proteins can be sensitive to temperature²¹, we tested the effect of temperature on the processing of CFTR Δ F508. Here we show that the processing of CFTR Δ F508 reverts towards that of wild-type as the incubation temperature is reduced. When the processing defect is corrected, cAMP-regulated Cl^- channels appear in the plasma membrane. These results reconcile previous contradictory observations and suggest that the mutant most commonly associated with cystic fibrosis is temperature-sensitive.

Figure 1a shows that wild-type CFTR (lanes 1 and 7) expressed in 3T3 fibroblasts at 37 °C is present in three forms, labelled bands A–C (ref. 13). Band A is unglycosylated protein; band B is core-glycosylated protein, which is sensitive to endoglycosidase H, a feature associated with processing in the endoplasmic reticulum; and band C is mature CFTR, which is more extensively glycosylated, a characteristic of processing in the Golgi complex. In contrast, CFTR Δ F508 expressed in cells at 37 °C is present predominantly in bands A and B (Fig. 1a, lane 2), suggesting that the protein does not reach the Golgi complex^{13,14}. Glycosylation *per se* does not alter the movement or function of CFTR (ref. 14, and D. N. Sheppard and M.J.W., unpublished), rather it is a marker of protein processing. As the conformation of nascent proteins and their subsequent processing and transport can be sensitive to temperature^{21,22}, we investi-

gated whether maturation of CFTR Δ F508 is temperature-sensitive.

When we grew the cells at lower temperatures for two days, we saw a shift in the glycosylation pattern of CFTR Δ F508 (Fig. 1a, lanes 3–5) towards that of more mature protein (band C). In contrast, the glycosylation pattern of wild-type CFTR did not change (Fig. 1b, lanes 2–4). Figure 1c shows this graphically:

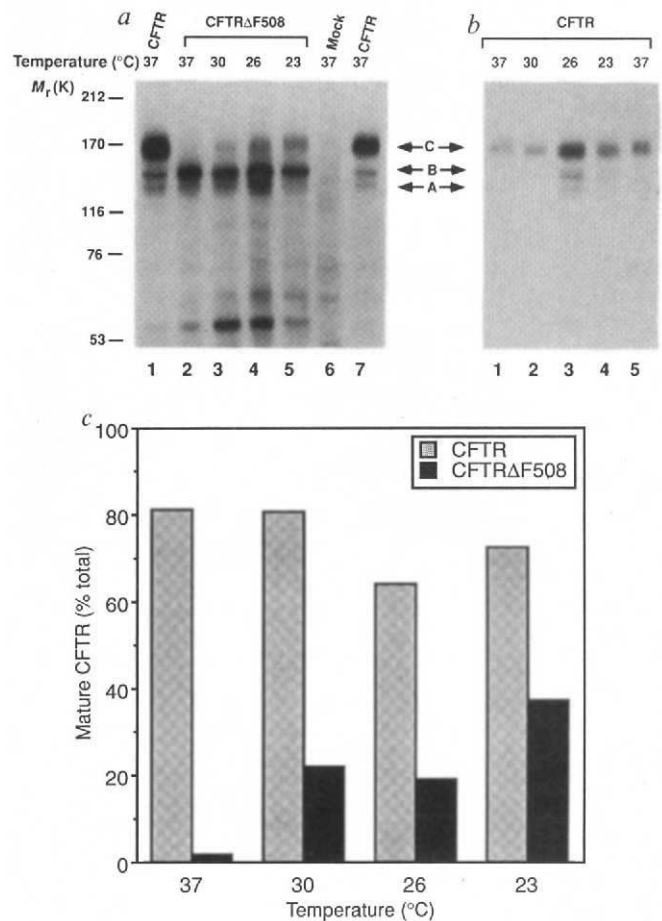


FIG. 1 Effect of temperature on processing of CFTR Δ F508 and wild-type CFTR. *a* and *b*, Autoradiographs of immunoprecipitated proteins, labelled by phosphorylation with [γ -³²P]ATP and the catalytic subunit of cAMP-dependent PKA, separated on 6% SDS–polyacrylamide gels. Cultures of 3T3 fibroblasts expressing CFTR Δ F508 (*a*) or CFTR (*b*) were grown at the indicated temperatures for 2 days. In *a*, control lanes represent wild-type CFTR (lanes 1 and 7) or mock-transfected (lane 6) cells grown at 37 °C. Bands A, B and C represent unglycosylated, core-glycosylated and fully glycosylated CFTR, respectively. *c*, Effect of temperature on the percentage of total CFTR that is present in the mature form. Values are calculated from radioactivity in band C divided by the sum of radioactivity in bands A, B and C, multiplied by 100; data are from panels *a* and *b*. Similar results were obtained in three other experiments.

METHODS. 3T3 fibroblasts expressing CFTR and CFTR Δ F508 (gift from S. Thompson and R. Mulligan) were generated as described⁴. Cells were seeded on 100-mm culture dishes and grown to 50–70% confluence before being shifted to the indicated temperature. At the end of the incubation period, cultures were washed twice with phosphate-buffered saline and solubilized for 30 min at 4 °C with lysis buffer (150 mM NaCl, 50 mM Tris, pH 7.5, aprotinin at 1 $\mu\text{g ml}^{-1}$, and 1 mM PMSF) containing 1% recrystallized digitonin. Equal amounts of starting protein (1.2–1.4 mg per sample) were immunoprecipitated with monoclonal antibody M13-1 and phosphorylated with [γ -³²P]ATP and the catalytic subunit of PKA as described¹³. Proteins were separated on 6% SDS–polyacrylamide gels and prepared for autoradiography. Quantitative assessment of radioactivity in gels and subtraction of background was done using a radioanalytic imaging system (AMBIS, San Diego). To avoid problems inherent in attempts to quantify absolute amounts of protein by immunoprecipitation, we present data as a percentage of total CFTR in band C.

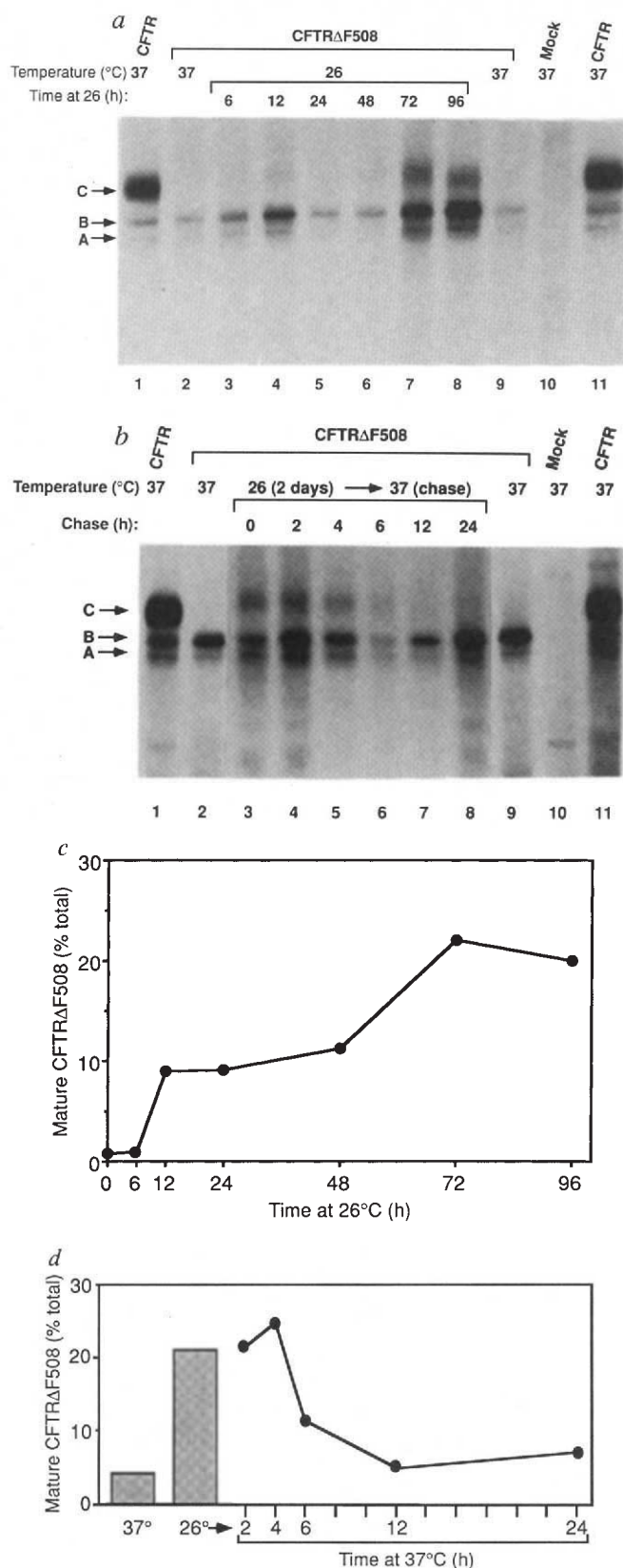


FIG. 2 Time-dependence and reversibility of the effect of incubation at a reduced temperature. *a*, Time course: 3T3 fibroblasts expressing CFTRΔF508 were grown for the indicated time at 26°C. An autoradiograph of a 6% SDS-polyacrylamide gel is shown. Similar results were obtained in two other experiments. *b*, Reversibility: 3T3 fibroblasts expressing CFTRΔF508 were grown for 2 d at 26°C and then grown at 37°C for the indicated time. An autoradiograph of a 6% SDS-polyacrylamide gel is shown. Similar results were obtained in two other experiments. *c* and *d*, Per cent of total CFTR radioactivity in band C; data are taken from *a* and *b*.

at decreased temperatures, the percentage of total CFTRΔF508 present as the mature, fully glycosylated band C increased, whereas the percentage of mature wild-type protein did not change appreciably. We found a similar effect of temperature on the glycosylation of CFTRΔF508 in C127 cells using a different expression system. This result suggests that at reduced temperatures, there is an increased flux of the mutant protein through the Golgi complex.

As we increased the length of incubation at a reduced temperature, the percentage of CFTRΔF508 present in the fully glycosylated form progressively increased (Fig. 2*a* and *c*). Moreover, the effect was reversible: when we first incubated the cells for two days at 26°C, and then returned the cultures to 37°C, the amount of mutant protein present as band C decreased with a half-life of ~7 h (Fig. 2*b* and *d*), a value similar to that observed for wild-type CFTR¹³. Presumably under these conditions the mature CFTRΔF508 produced at 26°C turns over and is not replaced by maturation of new protein.

If processing of the nascent protein is influenced by temperature, then mutant protein synthesized at 37°C should be processed normally during subsequent incubation at 26°C but not at 37°C. To test this possibility, we did the pulse-chase experiment shown in Fig. 3. CFTRΔF508 labelled with ³⁵S-methionine at 37°C (lane 1) could not be chased into the mature form at 37°C (lanes 2 and 3). But when CFTRΔF508 was labelled

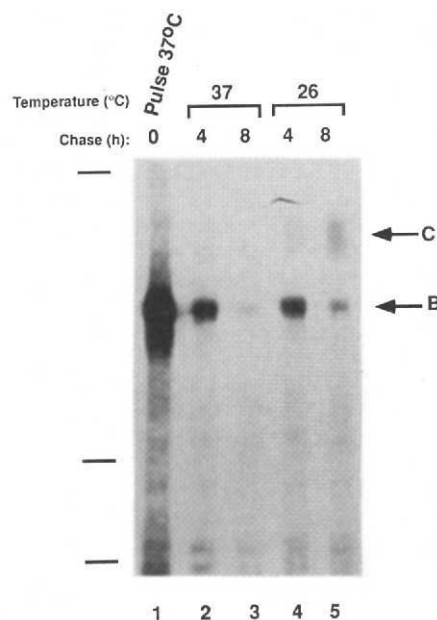


FIG. 3 Effect of temperature on processing efficiency of CFTRΔF508. C127 cells expressing CFTRΔF508 were grown for 40 h at 26°C, pulsed with 360 μCi [³⁵S]methionine per ml for 15 min at 37°C, then chased at the indicated temperature and time. Immunoprecipitates of cell lysates were separated on a 6% SDS-polyacrylamide gel, fluorographed with salicylate, and visualized by autoradiography. Bars at left indicate the migration of *M*_r markers of 200, 92.5 and 69K (from top to bottom). The processing efficiency of wild-type CFTR in C127 cells was unaffected by temperature (not shown). Similar results were obtained for cells grown at 37°C (not shown). **METHODS.** C127i cells (a cell line of mouse mammary epithelial origin) expressing CFTR or CFTRΔF508 were generated by calcium phosphate-mediated transfection with a bovine papilloma virus-based vector containing CFTR cDNA under control of the mouse metallothionein MT1 promoter and a neomycin-resistance gene (as the selectable marker) under the control of another copy of the MT1 promoter. Pulse-chase experiments were performed and assayed as described¹³, except that CDIM (carbon dioxide-independent medium; Gibco) was used for the chase, and the monoclonal antibody (24-1) was directed against a carboxy-terminal epitope (gift from P. Rafter). Under our conditions (100 μl of hybridoma supernatant and an overnight incubation) CFTR was maximally immunoadsorbed, so no additional CFTR was recovered after immunoprecipitation with monoclonal antibody M13-1 following immunoprecipitation with 24-1.

at 37 °C and the chase was then performed at 26 °C (lanes 4 and 5), some of the mutant protein was converted to the fully glycosylated form¹³. These results also show that correct processing of the mutant is not the result of an effect of temperature on protein synthesis.

These results indicated that at a reduced temperature, CFTR Δ F508 was delivered to the plasma membrane, where it could mediate Cl⁻ transport. To test this hypothesis, we used the whole-cell patch-clamp technique to measure cAMP-activated currents in the plasma membrane of cells expressing CFTR Δ F508 (all patch-clamping was at 35 °C). As we reported

previously^{16,17}, cAMP agonists failed to stimulate Cl⁻ currents in cells cultured at 37 °C (Fig. 4a). But after incubation at 30 °C for 2 days, cAMP agonists stimulated large chloride-selective currents. The same results were obtained in C127 cells expressing CFTR Δ F508 (Fig. 4a). Just as with the production of band C, when the temperature was returned to 37 °C for 24 h, the size of the cAMP-stimulated current decreased. Figure 4b and c shows that the cAMP-stimulated Cl⁻ current has a linear current-voltage relationship, that the channels are selective for Cl⁻ over Na⁺, and that there are little time-dependent voltage effects. These properties are similar to those of wild-type channels^{4,17,23}.

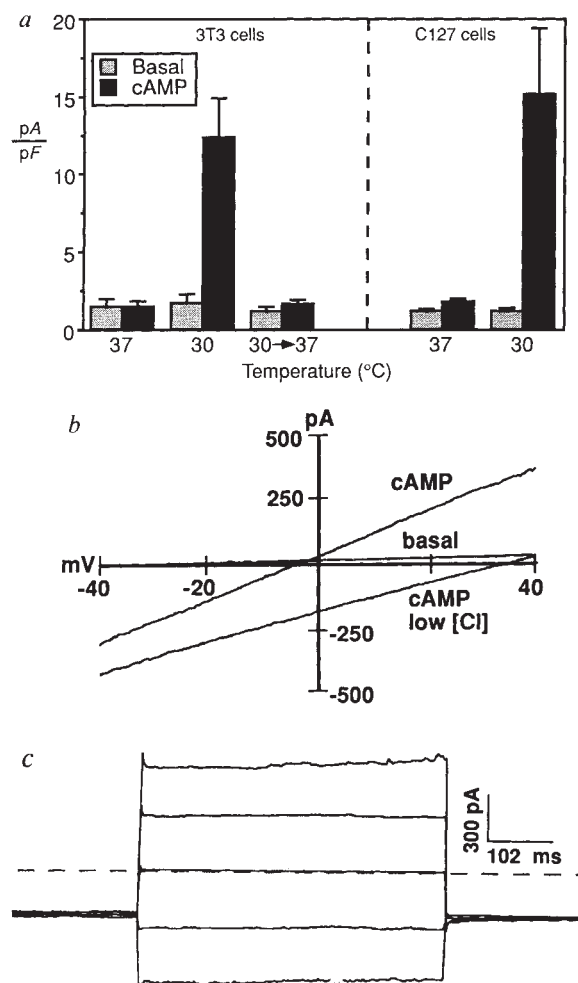


FIG. 4 Effect of temperature on cAMP-regulated Cl⁻ channel current. **a**, Effect of temperature on basal and cAMP-stimulated whole-cell currents. 3T3 fibroblasts (left panel) and C127 cells (right) expressing CFTR Δ F508 were cultured for 2 d at the indicated temperatures. One group of 3T3 fibroblasts was incubated at 30 °C for 2 d and then at 37 °C for 1 d. cAMP was increased by addition of 250 μ M 8-(4-chlorophenylthio)adenosine cyclic monophosphate (CPT-cAMP) and 20 μ M forskolin. The number of cells studied in each group and the cell capacitance was: 3T3 cells at 37 °C, 27.2 ± 1.2 pF ($n=12$); 3T3 cells at 30 °C, 30.4 ± 3.0 pF ($n=10$); 3T3 cells 30 °C and then 37 °C, 31.1 ± 8.8 ($n=5$); C127 cells at 37 °C, 27.6 ± 2.4 ($n=11$); C127 cells at 30 °C, 30.2 ± 3.8 ($n=10$). The cAMP-stimulated current was greater in cells cultured at 30 °C than at 37 °C for both 3T3 and C127 cells ($P < 0.012$). **b**, Current-voltage relationship of whole-cell currents under basal conditions after stimulation with cAMP agonists, and when the Cl⁻ concentration in the bathing solution was reduced from 140 mM to 22 mM by substitution of Cl⁻ with aspartate (low Cl⁻). Current-voltage relationships were obtained by a ramp of voltage from -40 mV to +40 mV. **c**, Representative whole-cell currents generated by voltage steps from -100 mV to +100 mV in 50-mV increments; holding voltage was -40 mV. **d**, Example of CFTR Cl⁻ channels in an excised, inside-out patch of membrane from 3T3 fibroblasts expressing CFTR Δ F508. Cell was incubated at 30 °C

for 2 d. Tracings were obtained under basal conditions, in the presence of 1 mM ATP, in the presence of 1 mM ATP plus 75 nM catalytic subunit of PKA, and after ATP and PKA were removed (wash). All traces were obtained at -100 mV. The patch contained multiple channels; dashed line indicates zero current level and channel opening is indicated by downward deflections. All patch-clamp studies were performed at 35 °C.

METHODS. Whole-cell and single-channel patch-clamp studies were done as described^{14,17}. For whole-cell studies the pipette (internal solution) contained (in mM) 120 N-methyl-D-glucamine (NMDG), 5 HEPES, 3 MgCl₂ and 1 Cs-EGTA (adjusted to pH 7.3 with HCl). The bath contained 140 NaCl, 10 HEPES, 1.2 CaSO₄, 1.2 MgSO₄, 10 dextrose (to pH 7.3 with NaOH). When bath Cl⁻ concentration was reduced, 120 sodium aspartate replaced 120 NaCl. For studies in excised inside-out patches, the pipette (external) solution contained (mM): 140 NMDG, 2 MgCl₂, 5 CaCl₂, 100 L-aspartic acid, and 10 HEPES (to pH 7.3 with HCl) (Cl⁻ concentration, 49 mM). The bath (internal) solution contained (mM): 140 NMDG, 3 MgCl₂, 1 Cs-EGTA, and 10 HEPES (to pH 7.3 with HCl) (Cl⁻ concentration, 147 mM). All experiments were done at 35 °C. Open-state probability (P_o) and single-channel conductance were determined from amplitude histograms. P_o was measured in patches containing ≤ 5 channels. The number of channels was determined from the maximum number simultaneously open with 3 mM ATP.

To confirm that the current results from expression of CFTR Δ F508 and to determine whether the Δ F508 mutation alters the properties of CFTR, we measured single-channel currents using excised, inside-out patches of membrane. Figure 4d shows that, just as with wild-type CFTR^{3,4,7}, there is little basal channel activity in excised patches. Addition of ATP to the cytosolic surface caused no activation, but a combination of ATP plus the catalytic subunit of cAMP-dependent protein kinase (protein kinase A; PKA) activated multiple channels. The current-voltage relationship was linear and the single-channel current at -100 mV was -1.1 ± 0.0 pA ($n = 6$), compared to -1.1 ± 0.0 pA ($n = 7$) for wild-type channels under similar conditions. When ATP and PKA were removed, activity returned to basal values. Despite these apparently normal biophysical properties and the normal regulation by cAMP (in the whole cell) and PKA (in excised patches), the channels had a reduced probability of being in the open state (P_o) at 1 mM ATP: P_o was 0.34 ± 0.02 ($n = 14$) for wild-type CFTR and 0.13 ± 0.01 ($n = 4$) for CFTR Δ F508. The reduced P_o of CFTR Δ F508 is similar to the value previously reported¹⁹.

Our findings reconcile previously incongruent observations. Earlier, we had failed to observe cAMP-regulated Cl^- channels when we expressed CFTR Δ F508 in airway epithelia and in heterologous cells^{16,17}. Moreover, several reports indicated that the apical membrane of cystic fibrosis-affected epithelia lacks a cAMP-regulated Cl^- conductance (reviewed in ref. 12). Yet CFTR Δ F508 could generate cAMP-regulated Cl^- channels when expressed in *Xenopus* oocytes¹⁸, Vero cells¹⁹, and Sf9 insect cells²⁰. Our data suggest that the reduced temperature used to culture *Xenopus* oocytes and Sf9 insect cells may be at least partly responsible for the detection of the functional protein in the plasma membrane. Although functional channels were detected in Vero cells grown at 37 °C, CFTR Δ F508 was produced by a very high-level expression system¹⁹. The data indicated that there might be a defect in protein transport¹⁹, so under those expression conditions the channel function could have represented a small fraction of the mutant protein that escaped the cellular quality control mechanisms to reach the plasma membrane, where it could generate a significant cAMP-stimulated Cl^- current.

Our results suggest that CFTR Δ F508 is a temperature-sensitive mutant. Temperature sensitivity of protein processing has been observed in other systems. For example, temperature-sensitive mutants of the vesicular stomatitis virus G protein have shown that oligomerization and movement in the cell can be promoted by reduced temperature²¹.

What are the implications for therapy? It is difficult to imagine any clinically acceptable therapeutic approach based on significant, repeated reductions in whole-body temperature. Nevertheless, the major cause of morbidity and mortality in cystic fibrosis is lung disease, and inhalation of cold air during hyperventilation can reduce airway temperature to values approaching 30 °C (refs 24, 25). Thus, it is possible that the processing of mutant CFTR might be altered *in vivo* by a reduction of temperature. Irrespective of the feasibility of such a manoeuvre, our results suggest that therapy designed to deliver mutant protein to the plasma membrane may be possible. □

14. Gregory, R. J. *et al. Molec. cell. Biol.* **11**, 3886-3893 (1991).
15. Denning, G. M., Ostedgaard, L. S. & Welsh, M. J. *J. Cell Biol.* **118**, 551-559 (1992).
16. Rich, D. P. *et al. Nature* **347**, 358-363 (1990).
17. Anderson, M. P., Rich, D. R., Gregory, R. J., Smith, A. E. & Welsh, M. J. *Science* **251**, 679-682 (1991).
18. Drumm, M. L. *et al. Science* **254**, 1797-1799 (1991).
19. Dalemans, W. *et al. Nature* **354**, 524-528 (1991).
20. Bear, C. E., Jensen, T. J. & Riordan, J. R. *Biophys. J.* **61**, A127 (1992).
21. Machamer, C. E. & Rose, J. K. *J. Biol. Chem.* **263**, 5955-5960 (1988).
22. Ljunggren, H.-G. *et al. Nature* **346**, 476-480 (1990).
23. Kartner, N. *et al. Cell* **64**, 681-691 (1991).
24. Gilbert, I. A., Fouke, J. A. & McFadden, E. R. Jr *J. appl. Physiol.* **63**, 1681-1691 (1987).
25. Deal, E. C. Jr, McFadden, E. R. Jr, Ingram, R. H. Jr & Jaeger, J. J. *J. appl. Physiol.* **46**, 484-490 (1979).

ACKNOWLEDGEMENTS. G.M.D. and M.P.A. contributed equally to this work. We thank D. Ries for helping with the immunoprecipitation and protein phosphorylation; L. DeBerg and P. Karp for culturing cells; T. Mayhew for typing the manuscript; and R. Gregory, Seng Cheng and our colleagues for discussion. This work was supported in part by the Howard Hughes Medical Institute and by grants from the National Heart, Lung and Blood Institute and the Cystic Fibrosis Foundation.

Predominant naturally processed peptides bound to HLA-DR1 are derived from MHC-related molecules and are heterogeneous in size

Roman M. Chicz, Robert G. Urban, William S. Lane*, Joan C. Gorga†, Lawrence J. Stern, Dario A. A. Vignali & Jack L. Strominger

Department of Biochemistry and Molecular Biology, and *Microchemistry Facility, Harvard University, Cambridge, Massachusetts 02138, USA

PEPTIDES bound to class I molecules are 8-10 amino acids long, and possess a binding motif representative of peptides that bind to a given class I allele¹⁻⁴. In the only published study of naturally processed peptides bound to class II molecules (mouse I-A^b and I-E^b), these peptides were longer (13-17 amino acids) and had heterogeneous carboxy terminals but precise amino-terminal truncations⁵. Here we report the characterization of acid-eluted peptides bound to HLA-DR1 by high-performance liquid chromatography, mass spectrometry and microsequencing analyses. The relative molecular masses of the peptides varied between 1,602 and 2,996 (13-25 residues), the most abundant individual M_r values being between 1,700 and 1,800, corresponding to an average peptide length of 15 residues. Complete sequence data were obtained for twenty peptides derived from five epitopes, of which all but one were from self proteins. These peptides represented sets nested at both the N- and C-terminal ends. Binding experiments confirmed that all of the isolated peptides had high affinity for the groove of DR1. Alignment of the peptides bound to HLA-DR1 and the sequences of 35 known HLA-DR1-binding peptides revealed a putative motif. Although peptides bound to class II molecules may have some related features (due to the nonpolymorphic HLA-DR α -chain), accounting for degenerate binding to different alleles⁶, particular amino acids in the HLA-DR β -chains presumably define allelic specificity of peptide binding.

The HLA-DR1 used in this study was solubilized by papain treatment to enable the material to be used both for crystallographic⁷ and bound-peptide analyses. The peptides bound to DR1 were acid-extracted and fractionated using microbore reversed-phase high-pressure liquid chromatography (RP-HPLC) (Fig. 1a). The absence of any detectable peptide material after a second extraction/RP-HPLC separation verified quantitative peptide extraction. Amino-acid analysis (ABI 420A/130A derivatizer/HPLC) of extracted peptide pools indicated that the yield was 70-80%, assuming total occupancy of purified DR1

Received 23 April; accepted 13 July 1992.

1. Anderson, M. P. *et al. Science* **253**, 202-205 (1991).
2. Bear, C. E. *et al. Cell* **68**, 809-818 (1992).
3. Tabcharani, J. A., Chang, X.-B., Riordan, J. R. & Hanrahan, J. W. *Nature* **352**, 628-631 (1991).
4. Berger, H. A. *et al. J. clin. Invest.* **88**, 1422-1431 (1991).
5. Rich, D. P. *et al. Science* **253**, 205-207 (1991).
6. Cheng, S. H. *et al. Cell* **66**, 1027-1036 (1991).
7. Anderson, M. P. *et al. Cell* **67**, 775-784 (1991).
8. Rommens, J. M. *et al. Science* **245**, 1059-1065 (1989).
9. Riordan, J. R. *et al. Science* **245**, 1066-1073 (1989).
10. Kerem, B. *et al. Science* **245**, 1073-1080 (1989).
11. Boat, T. F., Welsh, M. J. & Beaudet, A. L. in *The Metabolic Basis of Inherited Disease* Vol. 6 (eds Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D.) 2649-2680 (McGraw-Hill, New York, 1989).
12. Quinton, P. M. *FASEB J.* **4**, 2709-2717 (1990).
13. Cheng, S. H. *et al. Cell* **63**, 827-834 (1990).

† Present address: Division of Immunogenetics, Department of Pediatrics, University of Pittsburgh School of Medicine, Children's Hospital of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA.

TABLE 1 Naturally processed HLA-DR1-bound peptides

Protein source	Peptide sequence	Length	Residue	Peak	Calculated M_r	Observed M_r	Yield
HLA-A2	VGSDWRFLRGYHQYAYDG	18	103–120	5	2,190.4	2,190.4	39.5
	VGSDWRFLRGYHQYA	15	103–117	4	1,855.0	1,854.4	907.5
	VGSDWRFLRGYHQY	14	103–116	4	1,784.0	1,783.6	53.3
	GSDWRFLRGYHQYA	14	104–117	3	1,755.3	1,755.2	96.5
	SDWRFLRGYHQYA	13	105–117	3	1,698.2	1,698.8	48.8
							1,145.6
Invariant chain (Ii)	LRKPPKPVSKMRMATPLLQALPMG	25	96–120	9	2,733.5	2,734.5	40.5
	LPKPPKPVSKMRMATPLLQALPM	24	96–119	9	2,676.4	2,675.9	80.8
	PKPPKPVSKMRMATPLLQALPMG	24	97–120	7	2,620.2	2,619.7	91.5
	LPKPPKPVSKMRMATPLLQALP	23	96–118	7	2,545.2	2,544.5	112.2
	PKPPKPVSKMRMATPLLQALPM	23	97–119	8	2,563.2	2,562.3	145.0
	KPPKPVSKMRMATPLLQALPM	22	98–119	8	2,466.1	2,465.8	101.5
	PKPPKPVSKMRMATPLLQALP	22	97–118	6	2,432.0	2,431.7	72.5
	KPPKPVSKMRMATPLLQALP	21	98–118	6	2,334.9	2,334.2	31.6
	PPKPVSKMRMATPLLQALP	20	99–118	7	2,206.7	2,207.4	89.8
	KMRMATPLLQALPM	15	105–119	9	1,732.2	1,731.9	178.5
	KMRMATPLLQALP	14	105–118	7	1,601.0	1,600.2	162.0
							1,105.9
							48.8
(Na ⁺ + K ⁺)ATPase	IPADLRISANGCKVDNS	18	199–216	3	1,886.6	1,885.5	48.8
							48.8
Transferrin receptor	RVEYHFLSPYVSPKESP	17	680–696	4	2,035.3	2,036.8	30.3
							30.3
Bovine fetuin	YKHTLNQIDSVKVVPRRPT	19	56–74	1	2,338.7	2,338.5	32.5
	YKHTLNQIDSVKVVPRRP	18	56–73	2	2,237.6	2,236.5	69.0
							101.5

Amino-acid sequences (single-letter code) and mass determinations are shown for peptides isolated from HLA-DR1. Peptides were isolated as described for Fig. 1 and sequenced using an ABI 477A protein sequencer⁸. Collected fractions were as labelled in Fig. 1a. The peptide M_r s were calculated from the amino-acid sequences. Mass spectra were collected on a Finnigan-MAT (San Jose, California) TSQ 700 triple quadrupole mass spectrometer equipped with an electrospray ion source. The electrospray needle was operated with a voltage differential of 3 to 4 keV using a 50% methanol/0.5% acetic acid sample mixture delivered at 1 $\mu\text{l min}^{-1}$ by infusion injection. Observed M_r s were determined from the resulting spectra by computer-aided deconvolution of multiply charged m/z peaks³⁷. Yield refers to initial sequence levels of individual peptides measured in pmol, with the total yield of peptides from a single protein source summed and listed below the individual peptide yields.

with a molar equivalent of bound peptides corresponding to the size distribution determined by mass spectrometry (Fig. 1b). The RP-HPLC profiles obtained from DR1 extractions of multiple independent preparations were reproducible and profiles from either detergent-soluble or papain-solubilized DR1 were equivalent. To confirm that the peptides were in fact identical in detergent-soluble and papain-digested DR1, mass spectrometry and Edman sequencing analyses were done and revealed identical masses and sequences for analogous fractions from the two preparations.

Matrix-assisted laser desorption mass spectrometry (MALD-MS) was used to identify 111 species of unique mass in the eluted peptide pool of DR1 with an average size of 18 and a mode of 15 residues (Fig. 1b). Over 500 additional mass species in the M_r range of 13–25 residues were detected; however, the signal was too weak to assign individual masses with confidence. Multiple species of varying mass were detected in fractions corresponding to single RP-HPLC peaks indicating coelution of peptides. To characterize these peptides further, samples were analysed in parallel on a triple quadrupole mass spectrometer equipped with an electrospray ion source and by automated Edman degradation microsequencing⁸. Combining these two techniques allows crucial verification of both the N- and C-terminal amino acids of peptides contained in single fractions. The sequence and mass data acquired for 20 peptides isolated from DR1 are listed in Table 1. All the identified peptides aligned with complete identity to regions of proteins stored in the SWISS-PROT database⁹.

Of the 20 sequenced DR1-bound peptides, 16 were derived from the self proteins HLA-A2 and class II-associated invariant chain (Ii), which represented at least 26% of the total extracted peptide mass. These isolated peptides varied in length and were

truncated at both the N and C termini, suggesting that: (1) antigen processing occurs from both ends after binding to DR1; or (2) class II molecules bind antigen from a pool of randomly generated peptides. The yields from the peptide microsequencing indicated that HLA-A2 (Fig. 1a, peaks 3–5) and Ii (peaks 6–9) each represent at least 13% of the total DR1-bound peptides. All recognition of peptides derived from the class I molecule HLA-B7 presented by the class II molecule HLA-DR11.1 indicates that presentation of class I-derived peptides by class II may be a common phenomenon¹⁰.

The other four peptides identified were derived from two self proteins, transferrin receptor and the (Na⁺ + K⁺)ATPase, and one exogenous protein, bovine serum fetuin (Table 1). Each of these peptides occupied only 0.3–0.6% of the total DR1 population, significantly less than either the HLA-A2 or Ii peptides. Several groups have demonstrated that class II molecules *en route* to the cell surface intersect the pathway of incoming endocytic vesicles^{11–13}. Both recycling membrane proteins and endocytosed exogenous protein travel this common pathway. Hence the HLA-A2, transferrin receptor, (Na⁺ + K⁺)ATPase and bovine fetuin-derived peptides would all encounter DR1 in a similar manner. Ii associates with nascent class II molecules in the endoplasmic reticulum¹⁴, preventing antigen binding until the class II/Ii complex arrives at an endocytic compartment¹⁵, where Ii undergoes proteolysis^{16,17}, thus allowing peptide binding to proceed. Presumably the Ii peptides bound to DR1 were generated at this step.

Synthetic peptides corresponding to five of the peptides shown in Table 1 were made and characterized for their relative binding affinity to DR1. The influenza A haemagglutinin peptide (HA) 307–319 has been previously described as a high-affinity, HLA-DR1-restricted peptide^{18,19} and was thus chosen as the control

peptide. 'Empty' DR1 purified from recombinant insect cells was used in the binding experiments because of its higher binding capacity and 10-fold faster association kinetics than DR1 isolated from human cells²⁰. All the synthetic peptides were found to compete well (inhibition constant, $K_i < 100$ nM) against the HA peptide (Table 2). At first approximation, the Ii 106–119 peptide had the highest affinity of all the competitor peptides measured, equivalent to that determined for the control HA peptide. In addition to the K_i determinations, these peptides were able to confer resistance to SDS-induced chain dissociation of empty DR1 when analysed by SDS–polyacrylamide gel electrophoresis (PAGE), indicative of stable peptide binding^{21–23}. Neither of the two control peptides, β_2 -microglobulin 52–64 nor Ii 96–110, was able either to confer resistance to SDS-induced chain dissociation of DR1 or compete with HA 307–319 for binding to DR1; both of these peptides lack the putative binding motif reported here.

A putative DR1-binding motif based on the sequence alignments of the core epitopes (the minimum length) of naturally processed peptides is shown in Table 3. The key residues proposed in the motif are: a positively charged group is located at the first position referred to here as the index position for orientation (I); a hydrogen-bond donor at I+5; and a hydrophobic residue at I+9. A hydrophobic residue is also often found at I+1 and/or I-1. Every naturally processed peptide

sequenced from DR1 conforms to this motif (with the exception of the HLA-A2 peptide 103–116, which lacks residue I+9). Because the putative motif is not placed in a defined position with respect to the first amino acid and because of the irregular length of bound peptides, it is impossible to deduce a motif from sequencing of peptide pools, as was done for class I

TABLE 2 Peptide binding to HLA-DR1

Peptide	Length (amino-acid residues)	K_i against HA ⁺ 307–319 (nM)	SDS resistance
HLA-A2 103–117	15	49 ± 3	+
Ii 105–119	15	<10	+
Ii 96–119	24	33 ± 5	+
(Na ⁺ + K ⁺)ATPase 199–216	18	68 ± 9	+
Transferrin receptor 680–696	17	<10	+
Bovine fetuin 56–74	19	66 ± 18	+
HA 307–319	14	<10	+
Ii 96–110	15	>10 ⁴	–
β_2 m 52–64	13	>10 ⁴	–

The first six entries correspond to peptides found associated with HLA-DR1 and the sequences are shown in Table 1. Two control peptides were also tested: β_2 m 52–64 (SDLSFSKDWSEFYI) is from human β_2 -microglobulin, and Ii 96–110 (LPKPPKPVSKMRMAT) is a truncated version of the longest invariant chain-derived peptide isolated from HLA-DR1. Peptides were synthesized using solid-phase Fmoc chemistry, deprotected and cleaved using standard methods, then purified by RP-HPLC. Purified peptides were analysed by mass spectrometry and concentrations were determined by quantitative ninhydrin analysis. Inhibition constants (K_i) were measured as the concentration of test peptide that inhibited 50% of the ¹²⁵I-labelled HA 307–319 binding to empty HLA-DR1 produced in Sf9 insect cells²⁰. HA 307–319 was labelled using Na¹²⁵I and chloramine-T and isolated by gel filtration. Specific activity, determined by BCA assay (Pierce) and gamma counting, was 26,000 c.p.m. pmol⁻¹. Labelled peptide (10 nM) and 10 nM purified HLA-DR1 were mixed with 10 different concentrations (10 nM to 10 μ M) of synthetic cold competitor peptide in phosphate-buffered saline, pH 7.2, containing 1 mM EDTA, 1 mM PMSF, 0.1 mM iodoacetamide, and 3 mM Na₂S₂O₈, and incubated at 37 °C for 85 h. Free and bound peptide were separated by native gel electrophoresis³³ and bound radioactivity was quantitated using a Fujix imaging plate analyser (BAS 2000) after 4-h exposures on the phosphorimaging plates. Per cent inhibition was calculated as the ratio of background-corrected radioactivity in the sample to background-corrected radioactivity in a parallel sample containing no competitor peptide. Under these conditions, K_i measurements <10 nM could not be accurately determined. The ability of the synthetic peptides to confer resistance to SDS-induced chain dissociation of HLA-DR1 produced in insect cells was determined as described²⁰. Briefly, 20 μ M HLA-DR1 was incubated with fivefold excess of synthetic peptide at 37 °C for 85 h, in PBS (pH 7.2) with the protease inhibitor mixture described above. After incubation, the samples were analysed by SDS–PAGE with and without boiling before loading. Peptides that prevented SDS-induced chain dissociation are indicated by plus signs and those that did not by minuses.

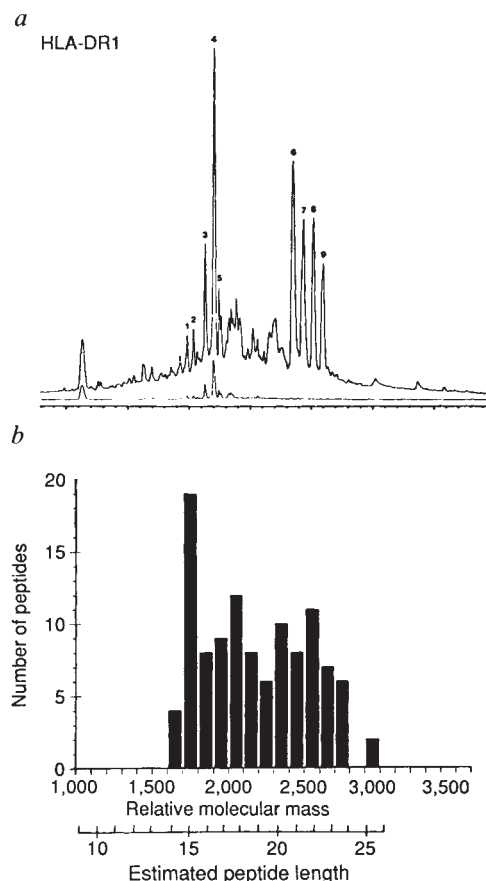


FIG. 1 Characterization of the peptide pools extracted from papain-digested HLA-DR1. *a*, The chromatogram represents the peptide repertoire as detected by ultraviolet absorbance. The ultraviolet absorbance for both 210 nm and 277 nm is shown at a full-scale absorbance of 0.500 with a retention window between 16 min and 90 min (each mark represents 2 min). *b*, Size distribution of isolated HLA-DR1-bound peptides determined by MALD-MS. The determined peptide masses in groups of 100 mass units were plotted against the number of isolated peptides identified by mass spectrometry. Peptide length was calculated by dividing the experimental mass by an average amino-acid M_r of 118. Although the weighted average for the length of bound peptide is ~18, a preference for DR1 binding of 15-amino-acid peptides is observed.

METHODS. DR1 was purified from the Epstein–Barr virus-transformed human B lymphoblastoid line LG-2. Cell growth, collection conditions and protein purification were as previously described^{7,36}. Protein concentration was determined by BCA assay (Pierce). All protein samples were next purified by size-exclusion HPLC on a TSK-250 column (300 × 7.8 mm) in 25 mM MES, 150 mM NaCl, pH 6.5, at a flow rate of 1 ml min⁻¹. DR1 (1 mg) was spin-concentrated to 50 μ l using a Centricon 10 microconcentrator. Acetic acid (1 ml, 10%) was added and the sample heated at 70 °C for 15 min then recentrifuged. The flow-through provided the acid-eluted peptide pool. An additional volume of acetic acid (1 ml, 10%) was added to the retentate and spin-concentrated to give a final acid-eluted volume of 2 ml. The peptide pool was concentrated to 50 μ l in a Savant Speed-Vac before HPLC separation. Peptides were separated on a Vydac microbore C-18 column as previously described³. Chromatographic analysis was monitored at multiple ultraviolet wavelengths to permit spectrophotometric evaluation of purified species before mass and sequence analyses. Collected fractions from the RP-HPLC separation of the HLA-DR1 peptide pool were analysed on a Finnigan-MAT LASERMAT MALD-MS using a sinapinic acid matrix, an accelerating voltage of 20 keV and a nitrogen laser (337 nm). Sample calibration was achieved by static fast-atom bombardment mass spectrometry of strong signal response samples.

TABLE 3 Putative HLA-DR1 peptide-binding motif

	Protein source	Peptide sequence	Length	Position	Ref.
(a)	HLA-A2	<u>SDWRFLRGYHQYA</u>	13	105-117	This study
	Invariant chain	<u>KMRMATPLLMQALP</u>	14	105-118	
	(Na ⁺ +K ⁺) ATPase	<u>IPADLRISANGCKVDNS</u>	18	199-216	
	Transferrin receptor	<u>RVEYHFLSPYVSPKESP</u>	17	680-696	
	Bovine fetuin	<u>YKHTLNQIDSVKVWPRRP</u>	18	56-73	
(b)	HEL	<u>KVFGRCELAAAMKRHGLD</u>	18	1-18	6
		<u>RNRCKGTDVQAWIRGCR</u>	18	112-129	6
	β_2m	<u>HPPHIEIQMLKNGKKI</u>	16	31-46	6
	PLA ₂	<u>NELGRFKHTDACRTH</u>	16	19-34	6
		<u>SKPKVYQWFDLRKY</u>	14	115-128	6
	NASE	<u>ATSTKKLHKEPATLKAIDG</u>	20	1-20	6
		<u>PATLKAIDGDTVKLMYKGQ</u>	20	11-30	6
		<u>DRVKLMYKGQPMFRLVD</u>	20	21-40	6
		<u>VAYVYKPNNTHEQHLRKSEA</u>	20	111-130	6
	HIV p13	<u>QKQEPIDKELYPLTSL</u>	16	97-112	6
	HIV p17	<u>GARASVLGGELDKWE</u>	16	1-16	6
	Influenza HA	<u>RTLYQNVTYVSVGTSTLNK</u>	20	187-206	6
	Influenza HA	<u>PKYVKQNTLKLAT</u>	13	307-319	17
	<i>P. falciparum</i> p190	<u>LKKLVFGYRKPLDNI</u>	15	249-263	25
	<i>P. falciparum</i> CS	<u>KHIEQYLKKIKNS</u>	13	329-341	27
	Chicken OVA	<u>DVFKELKVHANEINIF</u>	16	15-30	6
	DR1 β -chain	<u>GDTRPRFLWQLKFECHFFNG</u>	20	1-20	28
		<u>TERVRLLERCIYNQESVRFDS</u>	22	21-42	28
		<u>DLLEQRRAAVDTYCRHNYGVGESFT</u>	25	66-90	28
	p Cyt c	<u>KAERADLIAYLKQATAK</u>	17	88-104	6
	Myelin basic protein	<u>GRTQDENFVVHFFKNIVTPRTPPP</u>	24	75-98	6
(c)	Influenza matrix	<u>PLKAETIAQRLEDV</u>	13	19-31	6
	HIV p17	<u>RQILGQLQPSLQTGSE</u>	16	57-72	6
	β_2m	<u>IQVYSRHPPENGKPN</u>	16	7-22	6
	PLA ₂	<u>INTKCYKLEHPVTGGC</u>	16	85-100	6
	<i>P. falciparum</i> p190	<u>YKLNIFYDILLRAKL</u>	14	211-224	25
		<u>IDTLKKKNENIKEL</u>	13	338-350	25
	DR1 β -chain	<u>DVGEYRAVTELGPRDAEYWN</u>	20	43-62	28
	HIV p17	<u>ERFAVNPGELLETSGCC</u>	16	41-56	6
	HEL	<u>DNYRGYSLGNWVCAAKFESNFTQ</u>	23	20-42	6
	NASE	<u>EALVRQGLAKVAYVYKPNNT</u>	20	101-120	6
	HIV p25	<u>PIVQNLGQGMVHQAIS</u>	16	1-16	6
		<u>SALSEGATPQDLNTML</u>	16	41-56	6
	β_2m	<u>SFYILAHTFTPTTETD</u>	16	61-76	6
	PLA ₂	<u>KMYFNLINTKCYKLEH</u>	16	79-94	6

The putative HLA-DR1 peptide-binding motif includes three key amino acids; a positive charge at I, a hydrogen-bond donor at I+5, and a hydrophobic residue at I+9. The peptides are listed such that I, I+5, and I+9 are vertically aligned and underlined. *a*, The peptide core epitopes identified in Table 1; *b*, those peptides known to bind HLA-DR1 which totally conform to the described motif, and *c*, lists those peptides known to bind HLA-DR1 with either one key motif residue shifted or substituted at I by a large hydrophobic residue. These residues in *c* are in bold type.

molecules²⁴. The Ii 96-110 peptide has the I and I+5 motif residues in its sequence, but is missing eight additional amino acids found in Ii 105-118 (Table 3c).

A sequence comparison of 35 previously described DR1-binding synthetic peptides^{6,25-28} also supports this motif with 21 of the peptides (60%) having the precise motif, nine of the peptides (30%) containing a single shift at either I or I+9 and the remaining five (10%) having a single substitution at I (Table 3). In the latter, a positive charge at I is always replaced by a large hydrophobic residue (Table 3c); a pocket has been described in class I molecules that can accommodate this precise substitution²⁹. Contributions by the other eight amino acids in the motif or the length of the peptide have not been fully evaluated and may compensate for shifted/missing residues in those peptides that bind. Evaluation of the remaining 117 peptides that do not bind DR1 cited in those studies indicates that 99 (85%) of these peptides do not contain the proposed DR1 motif. Of the remaining 18 peptides (15%) containing the motif that do not bind to DR1, 6 (5%) bind to other DR alleles; the remaining 12 peptides may have unfavourable interactions at

other positions that interfere with binding. Because all HLA-DR alleles contain the same nonpolymorphic DR α -chain, the residues pointing inward from the α 1 helix and half of the floor of the groove are invariant³⁰. Thus some features of the motif for binding to DR1 might be expected to be shared by all DR alleles. Using T-cell recognition and/or binding of a limited number of analogue peptides, several studies have reported specific residues, primarily hydrophobic residues, at precise locations that are critical for peptide binding to DR1^{19,26,31-34}. The peptides described in those studies conform in part to the putative motif described herein. But some of the naturally processed peptides identified here do not fit these described motifs. Confirmation or rejection of this motif will await the refinement of the crystallographic structure of DR1 with bound peptide or studies with analogue peptides.

In contrast to the precise N-terminal cleavages observed in the previous study of six peptides bound to mouse I-A^b and five bound to I-E^b (ref. 5), the peptides bound to DR1 are heterogeneous at both the N and C termini. What might these data tell us about antigen processing in the class II pathway?

In contrast to peptides bound to class I molecules, which are predominantly nonamers¹⁻⁴, class II peptides are larger and display a high degree of heterogeneity both in length and the site of terminal truncation, implying that the mechanisms of processing for class I and class II peptides are substantially different. Furthermore, it may suggest that class II processing is a stochastic event and that a DR allele may bind peptides of different lengths from a complex random mixture. The heterogeneity observed may only be due to protection of bound peptides from further degradation. Thus, class II molecules would play an active role in antigen processing (as previously proposed³⁵) by protecting the bound peptides from complete degradation. Alternatively, the predominance of 15-amino-acid peptides bound to DR1 (as detected by both the MALD-MS and the yields of sequenced peptides) could be the result of trimming of bound peptides. In any event, the absence of peptides shorter than 13 and longer than 25 residues suggests that there are length constraints intrinsic to either the mechanism of peptide binding or antigen processing. In fact, a 15-residue peptide in an extended conformation can fit in an open cleft, but a longer peptide would have to extend beyond the cleft in such a conformation.

The predominance of peptides bound to DR1 that are derived from endogenously synthesized proteins, and particularly MHC-related proteins, may result from the evolution of a mechanism for presentation of self peptides in connection with the generation of self tolerance. The plethora of self peptides also has potentially important implications for the generation of autoimmune diseases linked to particular alleles of class II molecules, resulting from a breakdown of self tolerance.

Note added in proof: Another report describing naturally processed peptides extracted from murine class II molecules has been published³⁸. □

Received 21 April; accepted 9 July 1992.

1. Van Bleek, G. M. & Nathenson, S. G. *Nature* **348**, 213-216 (1990).
2. Rotzschke, O. et al. *Nature* **348**, 252-254 (1990).
3. Jardetzky, T. S., Lane, W. S., Robinson, R. A., Madden, D. R. & Wiley, D. C. *Nature* **353**, 326-329 (1991).
4. Hunt, D. F. et al. *Science* **255**, 1261-1263 (1992).
5. Rudensky, A. Y., Preston-Hurlburt, P., Hong, S.-C., Barlow, A. & Janeway, C. A. *Nature* **353**, 622-627 (1991).
6. O'Sullivan, D. et al. *J. Immunol.* **145**, 1799-1808 (1990).
7. Gorga, J. C., Brown, J. H., Jardetzky, T., Wiley, D. C. & Strominger, J. L. *Res. Immun.* **142**, 401-407 (1991).
8. Lane, W. S., Galat, A., Harding, M. W. & Schreiber, S. L. *J. Prot. Chem.* **10**, 151-160 (1991).
9. Pearson, W. R. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2444-2448 (1988).
10. Chen, B. P., Madrigal, A. & Parham, P. *J. exp. Med.* **172**, 779-783 (1990).
11. Cresswell, P. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8188-8192 (1985).
12. Neefjes, J. J., Stollorz, V., Peters, P. J., Geuze, H. J. & Ploegh, H. L. *Cell* **61**, 171-183 (1990).
13. Guagliardi, L. E. et al. *Nature* **343**, 133-139 (1990).
14. Jones, P. P., Murphy, D. B., Hewgill, D. & McDevitt, H. O. *Molec. Immun.* **16**, 51-60 (1978).
15. Roche, P. A. & Cresswell, P. *Nature* **345**, 615-618 (1990).
16. Thomas, L. J., Nguyen, Q. V., Elliott, W. L. & Humphreys, R. E. *J. Immunol.* **140**, 2670-2675 (1988).
17. Roche, P. A. & Cresswell, P. *Proc. natn. Acad. Sci. U.S.A.* **88**, 3150-3154 (1991).
18. Roche, P. A. & Cresswell, P. *J. Immunol.* **144**, 1849-1856 (1990).
19. Rothbard, J. B. et al. *Cell* **52**, 515-523 (1988).
20. Stern, L. J. & Wiley, D. C. *Cell* **68**, 465-477 (1992).
21. Sadegh-Nasseri, S. & Germain, R. N. *Nature* **353**, 167-170 (1991).
22. Dornmair, K., Rothenhauser, B. & McConnell, H. M. *Cold Spring Harb. Symp. quant. Biol.* **54**, 409-415 (1989).
23. Springer, T. A., Kaufman, J. F., Suddow, L. A., Mann, D. L. & Strominger, J. L. *J. biol. Chem.* **252**, 6201-6207 (1977).
24. Falk, K., Rotzschke, O., Stevanovic, S., Jung, G. & Rammensee, H. *Nature* **351**, 290-296 (1991).
25. Guttinger, M. et al. *Int. Immun.* **3**, 899-906 (1991).
26. Hill, C. M., Hayball, J. D., Allison, A. A. & Rothbard, J. B. *J. Immunol.* **147**, 189-197 (1991).
27. Guttinger, M. et al. *EMBO J.* **7**, 2555-2558 (1988).
28. Harris, P. E., Liu, Z. & Suciu-Foca, N. *J. Immunol.* **148**, 2169-2174 (1992).
29. Latron, F. et al. *Proc. natn. Acad. Sci. U.S.A.* **88**, 11325-11329 (1991).
30. Brown, J. H. et al. *Nature* **332**, 845-850 (1988).
31. Rothbard, J. B. & Taylor, W. R. *EMBO J.* **7**, 93-100 (1988).
32. Kropshofer, H., Bohlinger, I., Max, H. & Kalbacher, H. *Biochemistry* **30**, 9177-9187 (1991).
33. Jardetzky, T. S. et al. *EMBO J.* **9**, 1797-1803 (1990).
34. De Magistris, M. T. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 2990-2994 (1992).
35. Donermeyer, D. L. & Allen, P. M. *J. Immunol.* **142**, 1063-1068 (1989).
36. Gorga, J. C., Horejsi, V., Johnson, D. R., Raghupathy, R. & Strominger, J. L. *J. biol. Chem.* **262**, 16087-16094 (1987).
37. Hall, M. et al. in *Biological Mass Spectrometry* (eds Burlingame, A. L. & McCloskey, J. A.) 101-115 (Elsevier, Amsterdam, 1990).
38. Hunt, D. F. et al. *Science* **256**, 1817-1820 (1992).

ACKNOWLEDGEMENTS. We thank M. Gordy for sequence analysis of peptides. This work was supported by the Irvington Institute for Medical Research (R.M.C. and R.G.U.), the Medical Foundation/Charles A. King Trust Fund (J.C.G.), the Damon Runyon-Walter Winchell Cancer Fund (L.J.S.), the International Human Frontier Science Program Organization (D.A.A.V.) and the NIH (J.L.S.).

The structure of ribosomal protein S5 reveals sites of interaction with 16S rRNA

V. Ramakrishnan*† & Stephen W. White‡

* Biology Department, Brookhaven National Laboratory, Upton, New York 11973, USA

† MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

‡ Department of Microbiology, Duke University Medical Center, Durham, North Carolina 27710, USA

UNDERSTANDING the process whereby the ribosome translates the genetic code into protein molecules will ultimately require high-resolution structural information, and we report here the first crystal structure of a protein from the small ribosomal subunit. This protein, S5, has a molecular mass of 17,500 and is highly conserved in all lifeforms¹⁻⁴. The molecule contains two distinct α/β domains that have structural similarities to several other proteins that are components of ribonucleoprotein complexes. Mutations in S5 result in several phenotypes which suggest that S5 may have a role in translational fidelity and translocation. These include ribosome ambiguity or *ram*⁵, reversion from streptomycin dependence⁶ and resistance to spectinomycin⁶. Also, a cold-sensitive, spectinomycin-resistant mutant of S5 has been identified which is defective in initiation⁷. Here we show that these mutations map to two distinct regions of the molecule which seem to be sites of interaction with ribosomal RNA. A structure/function analysis of the molecule reveals discrepancies with current models^{8,9} of the 30S subunit.

The ribosomal protein S5 can be crystallized from the thermophilic organism *Bacillus stearothermophilus*¹⁰. The crystals are in the trigonal space group P3₂21 with cell dimensions $a = b = 59.3$ Å, $c = 109.8$ Å. We have cloned and overexpressed the gene for *B. stearothermophilus* S5 to provide sufficient material for structural analysis¹¹. The *B. stearothermophilus* protein is 55% homologous to its *Escherichia coli* counterpart^{1,2} and its function is almost certainly identical in the two organisms. We were able to determine the S5 structure with a single gold (KAu(CN)₂) derivative¹² by exploiting the variable wavelength properties of synchrotron radiation (see Fig. 1; details will be presented elsewhere).

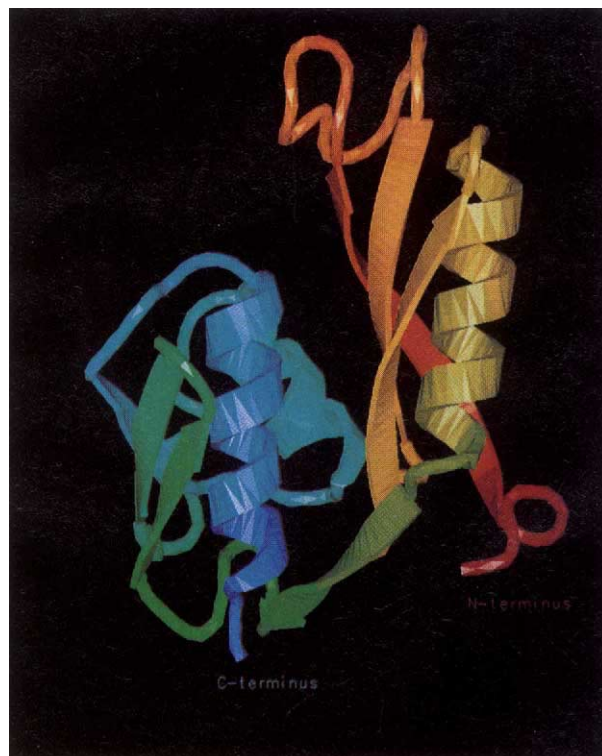
S5 has an α/β structure, and the polypeptide chain is folded into two distinct domains to create a somewhat elongated molecule (Fig. 2). There appears to be no flexibility between the two halves of the molecule, which are tightly associated by means of an extensive interdomain hydrophobic core. The N-terminal half consists of three contiguous β -strands arranged as an anti-parallel β -pleated sheet, and a 4-turn α -helix that lies on the sheet surface. The C-terminal half is similar in that it also contains a 3-strand antiparallel sheet, but the strands are not contiguous and there are two α -helices on one surface. The N and C termini are both disordered. The notation for the secondary structure elements and connecting loops, and their locations within the primary structure are shown in Fig. 3. A feature of the fold is that the two halves associate such that the empty surface of the β -sheet in domain 1 packs against the two α -helices of domain 2. This produces two unusual features. First, one surface of the β -sheet in domain 2 is totally exposed, and second, helices $\alpha 2$ and $\alpha 3$ are packed between two β -sheets.

S5 sequences are now known from 13 prokaryotic, eukaryotic and archaeobacterial organisms. The conservation of the S5 tertiary structure throughout evolution is evident from Fig. 3. The only major differences occur at the N and C termini, and in loop 3 which connects the two domains. The structurally important residues of S5 are particularly well conserved. These include hydrophobic amino acids at the core, and glycines and alanines

FIG. 1 Stereo electron density maps of S5 showing a cluster of 3 of the 4 phenylalanine residues in the molecule. The maps are contoured at the 1-standard-deviation level. *a*, The initial map that was used to trace the polypeptide chain and to build the molecule. Data were collected on beam-line X12C of the National Synchrotron Light Source (Brookhaven National Laboratory) using an Enraf-Nonius FAST area detector³⁰. Native data were collected by a rotation about the a^* axis, and 14,806 reflections were measured to give 5,829 unique reflections to 2.7 Å resolution with a merging R -factor of 5.1%. Derivative (KAu(CN)₂) data were collected about the a axis, which allowed the simultaneous collection of anomalous pairs with considerable redundancy. A total of 28,369 reflections were merged to give 6,020 unique reflections to 2.7 Å, with a merging R -factor of 7.3%. The X-ray wavelength (1.04 Å) was chosen to maximize the anomalous signal from the gold atoms. The final heavy-atom refinement values for R_{Cullis} and R_{Kraut} to 2.7 Å were 0.47 and 0.12 respectively, and the combined mean figure of merit for isomorphous and anomalous data was 0.77. The map was then solvent-flattened assuming a solvent volume fraction of 0.5 (less than the estimated 0.6 to avoid truncation of density). This improved the mean figure of merit to 0.89; the chain could be continuously traced, with the exception of three residues (28–30) that form part of a loop and are partially disordered. Data were processed by the program MADNES, phase and map calculations were performed by the package PHASES³¹, and the molecule was built using the program O (ref. 32). *b*, $2F_o - F_c$ map of the same region after refinement of the initial model with X-PLOR³³. Using all data between 6 Å and 2.8 Å with no water molecules added, the R factor is currently 22% with a deviation from ideality of 0.017 Å in bond lengths and 3.74° in bond angles.



FIG. 2 Ribbon diagram of S5 produced by the program O (ref. 32). The secondary structure elements are colour-ramped red to blue, N terminus to C terminus, to show the path of the polypeptide chain. The molecule has an α/β structure (see Fig. 3 for the secondary structure notation), and is folded into two distinct domains, each containing a 3-strand, antiparallel β -pleated sheet. The N-terminal domain (right) contains one α -helix, and the C-terminal domain (left), two α -helices; the domains associate such that the latter are sandwiched between the two β -sheets. This arrangement exposes one surface of the C-terminal β -sheet which is populated by several conserved hydrophobic residues. Note the close approach of the third β -strand and the α -helix (on the nearside of the molecule) in the N-terminal domain, and the resulting gap between the helix and the first β -strand (on the backside of the molecule). This feature necessitates alanine and glycine residues along the connecting surfaces of the first pair, and results in the exposure of the hydrophobic core between the second pair. The residues at the N and C termini (3 and 19 respectively) are disordered. The molecule is rather flat and elongated, with approximate dimensions 45 Å × 40 Å × 25 Å.



in positions where the fold juxtaposes regions of the polypeptide chain. Notable among the latter is the close approach between almost the entire lengths of β_3 and α_1 (Fig. 2).

The structures of three small proteins that are components of ribonucleoprotein complexes have been determined previously. These include two prokaryotic ribosomal proteins L12CTF (ref. 13) and L30 (ref. 14), and the RNA recognition motif A-RRM1 from the (human) U1snRNP^{15,16}. All three have similar structures involving helices packed against one side of an anti-parallel β -pleated sheet. The identical secondary structure topologies in

FIG. 3 Amino-acid sequence comparisons for ribosomal protein S5 (one-letter code). Also shown are the locations of the secondary structure elements deduced from the electron density map. The four sequences include representatives from each of the three principal life forms, two from prokaryotes: *Escherichia coli*¹ (COLI) and *Bacillus stearothermophilus*² (STEARO); one from archaeobacteria: *Halobacterium marismortui*³ (MARIS); and one eukaryote: *Saccharomyces cerevisiae*⁴ (YEAST). The numbering scheme is based on the *B. stearothermophilus* sequence. Only the region visible in the electron density map is shown; the N and C termini are disordered, of variable length and show little or no homology. The only variable region within the body of the molecule is loop 3 which connects the two domains. This loop is considerably longer in the eukaryotic and archaeobacterial molecules. Residues in the hydrophobic core are indicated in bold italics. Underlined residues are glycines, alanines and prolines in structurally restricted regions of the molecule. Circled residues are sites of mutation in *E. coli* that produce spectinomycin resistance (residues 20, 21, 22), and reversion from streptomycin dependence and ribosome ambiguity or *ram* (residues 104, 112). Boxed residues are conserved arginine and lysine residues in loop 2 that are proposed to interact with 16S rRNA.

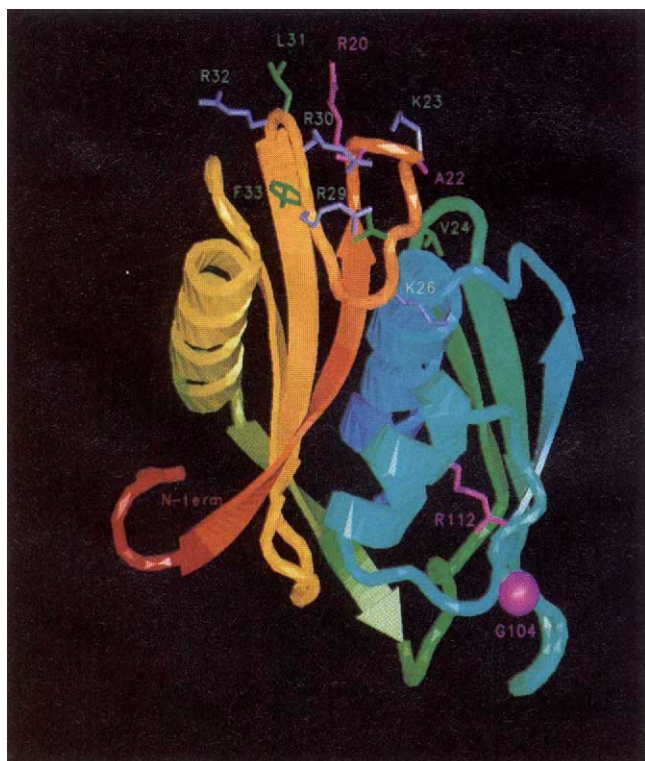
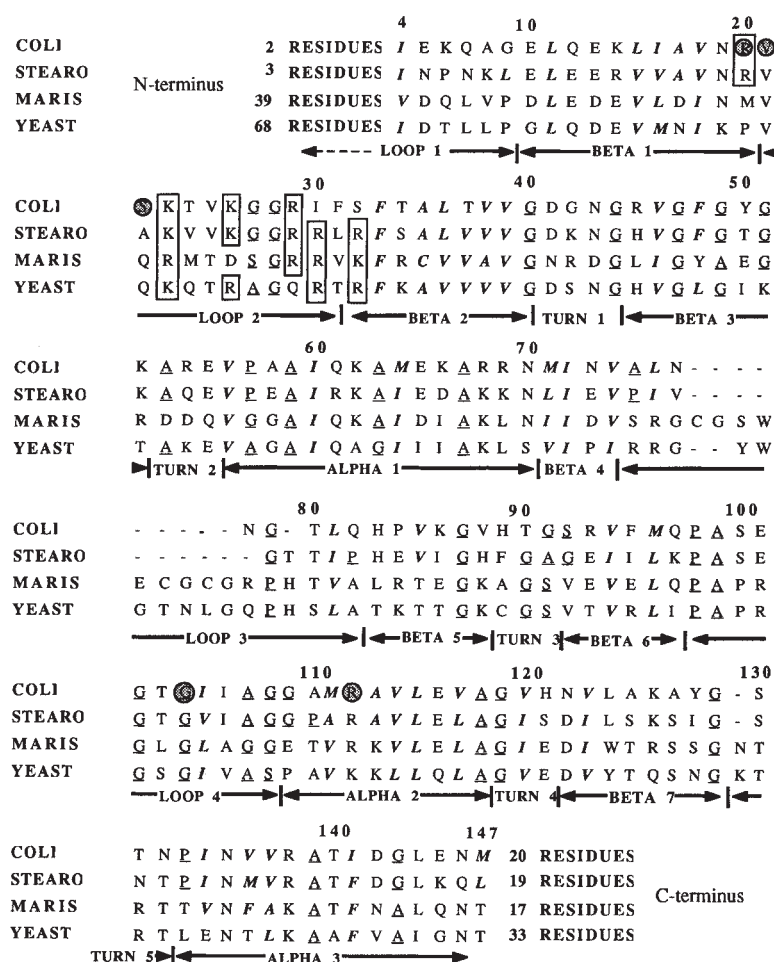


FIG. 4 Ribbon diagram of S5 produced by the program O (ref. 32) showing the two proposed sites of interaction with 16S rRNA. The secondary structure elements are colour-ramped red to blue, N terminus to C terminus, to show the path of the polypeptide chain. The N-terminal domain is to the left, the C-terminal domain is to the right, and the C-terminal helices are sandwiched between the two β -pleated sheets. The first site of interaction is on the C-terminal domain at the bottom right of the molecule. Two residues, glycine 104 and arginine 112 (magenta), are highly conserved, adjacent, and can be mutated to produce the *ram* and the revertant from streptomycin dependence phenotypes (see Fig. 3 and text). The second site is loop 2 (see Fig. 3 for notation) in the N-terminal domain at the top of the molecule centred on phenylalanine 33 (green). This contains several highly conserved arginine (20, 29, 30 and 32) and lysine (23 and 26) residues (blue), and residues (20, 21 and 22; magenta) that can be mutated to produce the spectinomycin-resistant phenotype (see Fig. 3 and text). Note that for reasons of clarity the side chain of residue 21 is not shown. Also shown are several hydrophobic residues (green) in this loop.

these molecules has led to the suggestion that they evolved from an ancient RNA-binding protein^{15,17}. This simple evolutionary scenario is complicated by S5. At first sight it would appear that each domain contains the typical helix-sheet structure and that the protein is a result of gene duplication from a more primitive member of this general family. But closer inspection reveals that the topological arrangements within each domain are not only different from each other but also from the other three proteins.

As mutations within S5 have important effects on translation (see earlier), the molecule represents an excellent probe of the structural basis of ribosome function. The mutations that alter

the ribosome's response to streptomycin and spectinomycin are particularly revealing. Streptomycin acts by reducing translational fidelity, whereas spectinomycin inhibits the translocation of peptidyl transfer RNA from the A site to the P site¹⁸. Both antibiotics interact directly with 16S rRNA; streptomycin with the base of the 900 stem-loop (helix 27; ref. 19) in the vicinity of nucleotide C912 (refs 18, 20), and spectinomycin with the helix 34 (ref. 19) close to the base-paired nucleotides G1,064–C1,192 (refs 18, 21). The mutations within S5 that relate to these antibiotics are clustered into two distinct regions of the molecule. The first involves residues glycine at position 104 and arginine 112, which are highly conserved and can be mutated to produce the streptomycin revertant and *ram* phenotypes¹ (Fig. 3). These are adjacent on the C-terminal half of the molecule close to the domain interface (Fig. 4). The second region is loop 2 (Fig. 4), which encompasses the amino acids responsible for spectinomycin resistance¹ (Fig. 3). This forms a mini-domain centred on phenylalanine 33 and contains six highly conserved arginine and lysine residues. As conserved arginine and lysine residues are often vital in protein–nucleic acid interactions²², we propose that the two S5 regions bind 16S rRNA close to the respective antibiotic-binding sites and that mutations directly modulate their mechanism of action.

How well does this scheme agree with current models for the organization of the 30S subunit^{8,9}? As regards the interaction with helix 27, the agreement is good. Both S5 and helix 27 are accepted components of the 'proofreading domain' that also contains S4, S12 and the 530 stem-loop (helix 18; ref. 19) and 5' terminus of 16S rRNA²³. These components are clustered together in the ribosome^{24–27} and are involved in translational fidelity. Mutations in this region of S5 also produce a *ram* phenotype⁵, in support of its association with the proof-reading components. The interaction with helix 34 is more problematical. In present models^{8,9}, these components are some 70 Å apart and communicate by an indirect 'allosteric' process involving S2 (ref. 23). We propose an alternative scheme which is consistent with both the biochemical data and the structure. Proteins S2 and S3 protect identical nucleotides in helix 34 (ref. 27). It has been suggested that these 'polyspecific' effects are not due to a direct association of the proteins with helix 34 but rather to a conformational change in the RNA²⁷. We suggest that this S2/S3-dependent change brings about an interaction between helix 34 and S5 which stabilizes the new and functional RNA conformation. Such a scheme would explain why S5 has no apparent effect on helix 34 in the absence of S2 and S3 (ref. 27), and is compatible with the relative locations of S2, S3 and S5 (refs 24, 25). The scheme is also consistent with the 30S assembly map which shows that the addition of S2 and S3 is dependent upon the presence of S5 (ref. 28). This assembly step may involve the binding of S2 to S5, which are adjacent in the particle^{24,25}; in fact, a hydrophobic region on S5 between $\beta 1$ and $\alpha 1$ is ideally positioned to be the contact site as it is close to loop 2 (Fig. 2). Finally, our results are in agreement with another report that strongly supports a revised 30S model in which S5 and helix 34 are neighbours²⁹. □

16. Nagai, K., Oubridge, C., Jessen, T. H., Li, J. & Evans, P. R. *Nature* **348**, 515–520 (1990).
17. Leijonmark, M. et al. *Proteins* **3**, 243–251 (1988).
18. Moazed, D. & Noller, H. F. *Nature* **327**, 389–394 (1987).
19. Brimacombe, R. *Biochimie* **73**, 927–936 (1991).
20. Montandon, P. E., Wagner, R. & Stutz, E. *EMBO J.* **5**, 3705–3708 (1986).
21. Sigmund, C., Ettayebi, M. & Morgan, E. A. *Nucleic Acids Res.* **12**, 4653–4663 (1984).
22. Ollis, D. L. & White, S. W. *Chem. Rev.* **87**, 981–995 (1987).
23. Noller, H. F. A. *Rev. Biochem.* **60**, 191–227 (1991).
24. Capel, M. S. et al. *Science* **238**, 1403–1406 (1987).
25. Lambert, J. M., Boileau, G., Cover, J. A. & Traut, R. R. *Biochemistry* **22**, 3913–3920 (1983).
26. Osswald, M. et al. *Nucleic Acids Res.* **15**, 3221–3240 (1987).
27. Stern, S., Powers, T., Changchien, L.-M. & Noller, H. F. *Science* **244**, 783–790 (1989).
28. Held, W. A., Mizushima, S. & Nomura, M. *J. Biol. Chem.* **248**, 5720–5730 (1973).
29. Dontsova, O., Kopylov, A. & Brimacombe, R. *EMBO J.* **10**, 2613–2620 (1991).
30. Arndt, U. W. *Meth. Enzym.* **114**, 472–485 (1985).
31. Furey, W. & Swaminathan, S. *Am. Crystallogr. Assn. Abstr. Ser. 2*, Vol. 18, 73 (1990).
32. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. *Acta Crystallogr.* **A47**, 110–119 (1991).
33. Brünger, A. T. *J. molec. Biol.* **203**, 803–816 (1988) (Program X-PLOR, Yale Univ., 1989).

ACKNOWLEDGEMENTS. This letter is dedicated to the memory of H.-G. Wittmann in whose laboratory the work was initiated. We thank R. Brimacombe, H. Noller and P. Wollenzien for discussion; R. M. Sweet and P. Lee for help with data collection; S. E. Gerchman and H. Kycia for technical assistance; W. K. Joklik and F. W. Studier for encouragement; and T. Leaf-Jones, P. McLaughlin and M. Kjeldgaard for help with model-building using the program O. This work was supported by a grant from the NIH, and by the Office of Health and Environmental Research of the US Department of Energy. One of us (V.R.) was the recipient of a Guggenheim Fellowship during part of this work.

Convergence of 9-*cis* retinoic acid and peroxisome proliferator signalling pathways through heterodimer formation of their receptors

Steven A. Kliewer*, Kazuhiko Umesono*,
Daniel J. Noonan†, Richard A. Heyman†
& Ronald M. Evans*

* Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, California 92037, USA

† Ligand Pharmaceuticals, 9393 Towne Centre Drive, Suite 100, San Diego, California 92121, USA

PEROXISOMES are cytoplasmic organelles which are important in mammals in modulation of lipid homeostasis, including the metabolism of long-chain fatty acids and conversion of cholesterol to bile salts (reviewed in refs 1 and 2). Amphipathic carboxylates such as clofibrate acid have been used in man as hypolipidaemic agents and in rodents they stimulate the proliferation of peroxisomes. These agents, termed peroxisome proliferators, and all-*trans* retinoic acid activate genes involved in peroxisomal-mediated β -oxidation of fatty acids^{1–4}. Here we show that the receptor activated by peroxisome proliferators⁵ and the retinoid X receptor- α (ref. 6) form a heterodimer that activates acyl-CoA oxidase gene expression in response to either clofibrate acid or the retinoid X receptor- α ligand, 9-*cis* retinoic acid, an all-*trans* retinoic acid metabolite^{7,8}; simultaneous exposure to both activators results in a synergistic induction of gene expression. These data demonstrate the coupling of the peroxisome proliferator and retinoid signalling pathways and provide evidence for a physiological role for 9-*cis* retinoic acid in modulating lipid metabolism.

The peroxisome proliferator responsive element (PPRE) located in the rat acyl-CoA oxidase (AOX) promoter is composed of two direct AGG(A/T)CA repeats separated by a single nucleotide (Fig. 1a)^{9–11} and thus conforms with previously described retinoid X response elements^{12,13}. Indeed, cotransfection of the retinoid X receptor- α (RXR α) expression plasmid in the presence of 9-*cis* retinoic acid (Fig. 1b, 9-*cis* RA, RXR) resulted in activation of a reporter construct containing the AOX promoter upstream of the luciferase gene (AOX-LUC). As previously shown⁹, the AOX promoter was also induced by the peroxisome proliferator-activated receptor (PPAR) in the presence of clofibrate acid (Fig. 1b, clofibrate acid, PPAR). Interest-

Received 6 May; accepted 15 July 1992.

1. Wittmann-Liebold, B. & Greuer, B. *FEBS Lett.* **95**, 91–98 (1978).
2. Kimura, M. *J. Biol. Chem.* **259**, 1051–1055 (1984).
3. Scholzen, T. & Arndt, E. *Molec. gen. Genet.* **228**, 70–80 (1991).
4. All-Robyn, J. A., Brown, N., Otake, E. & Liebman, S. W. *Molec. cell. Biol.* **10**, 6544–6553 (1990).
5. Piepersberg, W., Böck, A. & Wittmann, H.-G. *Molec. gen. Genet.* **140**, 91–100 (1975).
6. Piepersberg, W., Böck, A., Yaguchi, M. & Wittmann, H.-G. *Molec. gen. Genet.* **143**, 43–52 (1975).
7. Nomura, M. *Cold Spring Harb. Symp. quant. Biol.* **52**, 653–663 (1987).
8. Stern, S., Weiser, B. & Noller, H. F. *J. molec. Biol.* **204**, 447–481 (1988).
9. Brimacombe, R., Atmadja, J., Stiege, W. & Schüler, D. *J. molec. Biol.* **199**, 115–136 (1988).
10. Appelt, K., White, S. W. & Wilson, K. S. *J. Biol. Chem.* **258**, 13328–13330 (1983).
11. Ramakrishnan, V. & Gerchman, S. E. *J. Biol. Chem.* **266**, 880–885 (1991).
12. White, S. W., Appelt, K., Dijk, J. & Wilson, K. S. *FEBS Lett.* **163**, 73–75 (1983).
13. Leijonmark, M. & Liljas, A. *J. molec. Biol.* **195**, 555–580 (1987).
14. Wilson, K. S., Appelt, K., Badger, J., Tanaka, I. & White, S. W. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7251–7255 (1986).
15. Hoffmann, D. W., Query, C. C., Golden, B. L., White, S. W. & Keene, J. D. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2495–2499 (1991).

ingly, cotransfection of expression plasmids for both PPAR and RXR α in the presence of clofibric acid and 9-*cis* retinoic acid resulted in a synergistic increase in the activity of the AOX promoter (Fig. 1b, clofibric acid + 9-*cis* RA, PPAR + RXR).

To investigate whether the cooperativity between PPAR and RXR α occurred through the PPRE, cotransfection experiments were performed with a reporter construct containing three copies of the PPRE upstream of the thymidine kinase promoter fused to the luciferase gene (PPRE₃-TK-LUC). As expected, cotransfection of either PPAR or RXR α expression plasmids in the presence of clofibric acid or 9-*cis* retinoic acid, respectively, resulted in induction of reporter expression (Fig. 1c, clofibric acid or 9-*cis* RA). As in the case of the intact AOX promoter, coexpression of PPAR and RXR α in the presence of both activators resulted in a synergistic increase in PPRE₃-TK-LUC expression (Fig. 1c, clofibric acid + 9-*cis* RA). Thus, PPAR and RXR α cooperate in transactivating gene expression through the PPRE.

The synergism between PPAR and RXR α on the PPRE suggested the possibility of a heterodimeric interaction between the two receptors. To test this possibility, coimmunoprecipitation experiments were performed using bacterially expressed RXR α and *in vitro*-synthesized, [³⁵S]methionine-labelled PPAR. As shown in Fig. 2a, when mixed, polyclonal antiserum against RXR α efficiently coprecipitated PPAR, indicating the formation of a solution heterodimer. Using the PPRE as probe, gel mobility shift experiments were performed to examine the binding properties of the PPAR-RXR α complex. Neither PPAR nor RXR α alone bound efficiently to the radiolabelled PPRE (Fig. 2b, lanes 2 and 3). Mixing the PPAR and RXR α proteins, however, resulted in the appearance of a strong, shifted complex

(Fig. 2b, lane 4) which was further shifted by inclusion of the polyclonal antiserum (Fig. 2b, lane 5). Thus, PPAR and RXR α are capable of forming a complex in solution that binds DNA in a cooperative fashion.

RXR also forms heterodimers with the receptors for vitamin D, thyroid hormone and retinoic acid (VDR, TR and RAR, respectively), which bind cooperatively to their cognate hormone response elements (HREs)¹⁴⁻¹⁷. *In vivo*, the selectivity in the hormonal response is achieved through these receptor complexes recognizing direct repeats spaced by 3, 4 or 5 nucleotides, respectively (3-4-5 rule)¹⁸. The comparative binding specificity of the PPAR-RXR α complex was assessed by gel mobility shift experiments using as competitors a series of synthetic HREs composed of two AGGTCA direct repeats separated by spacers from zero to five nucleotides in length (DR 0-5; ref. 13). As predicted from *in vivo* studies¹⁸, heterodimers formed between RXR α and the VDR, TR or RAR interacted most strongly with DR-3, DR-4 and DR-5, respectively (Fig. 3a). In contrast, the PPAR-RXR α complex interacted with the DR-1 oligonucleotide only. The RAR-RXR α complex also bound to DR-1 (Fig. 3a) indicating that there is the potential for competitive interactions between these heterodimers.

Consistent with the observed DR-1 binding preference, the PPAR-RXR α complex interacted strongly with other natural DR-1-like HREs found in the 3-ketoacyl-CoA thiolase, chicken ovalbumin and cellular retinol-binding protein type II promoters as well as the PPRE (Fig. 3b, lanes 1-4). In contrast, the PPAR-RXR α complex bound only weakly to the retinoic acid response element of the RAR- β promoter and the thyroid hormone response element of the Moloney leukaemia virus long terminal repeat, and not at all to the osteopontin vitamin D

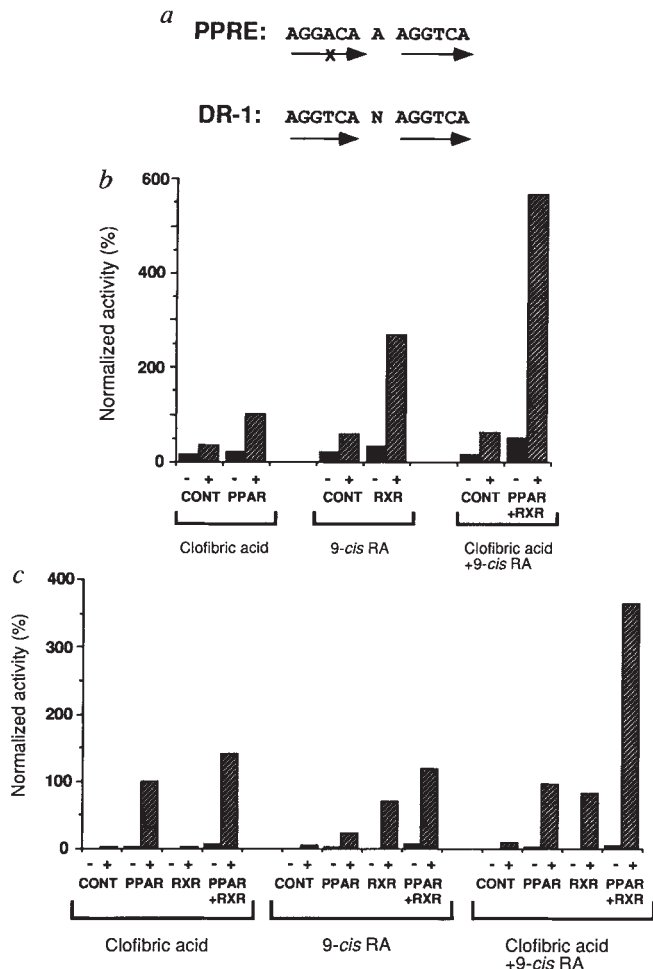


FIG. 1 PPAR and RXR α transactivate reporter expression cooperatively through the PPRE. **a**, The sequence of the PPRE located between nucleotides -558 and -570 of the AOX promoter is compared to that of the idealized DR-1 HRE. The positions of the AGGTCA direct repeats are indicated by arrows. X, nucleotide substitution in this motif. **b**, CV-1 cells were cotransfected with reporter construct AOX-LUC and either the control (CONT) Rous sarcoma-virus promoter-chloramphenicol acetyl transferase (RS-CAT) or expression plasmids RS-rat(r)PPAR (PPAR) and/or RS-human(h)RXR α (RXR) as indicated. Cells were subsequently treated with carrier dimethylsulphoxide (DMSO) (-) or hormones (+), including clofibric acid, 9-*cis* retinoic acid (RA), or clofibric acid and 9-*cis* RA as indicated. Luciferase activity is presented as per cent normalized response where induced PPAR activity in the presence of clofibric acid is arbitrarily set at 100%. **c**, CV-1 cells were cotransfected with reporter construct PPRE₃-TK-LUC and either the control RS-CAT or expression plasmids RS-rPPAR and/or RS-hRXR α as indicated.

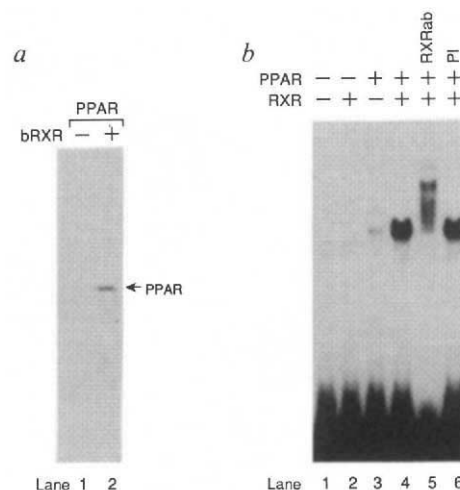
METHODS. Expression plasmid RS-rPPAR was constructed by inserting the rat PPAR cDNA (D.J.N., submitted) into the *KpnI/BamHI* sites of pRS (ref. 6). The RS-human RXR α expression plasmid has been described¹². The reporter plasmid AOX-LUC was constructed by amplification from rat genomic DNA of the 5' flanking sequences (nucleotides -602 to +20) of the rat AOX promoter by 30 cycles of the polymerase chain reaction (PCR) and directional insertion into *BamHI/XhoI*-cut pLUC (ref. 12). Oligonucleotides for PCR contained *BamHI* (5') and *XhoI* (3') restriction sites to facilitate subcloning. The reporter construct PPRE₃-TK-LUC was constructed by inserting 3 copies of an oligonucleotide encoding the PPRE (GTGACAGGGGACAGGACAAAGGTCACTCGGGAGTCGAC)⁹⁻¹¹ in direct orientation into the unique *SalI* site of the basal reporter construct TK-LUC (ref. 12). The orientation of the PPRES was confirmed by sequencing. Transfection assays were performed using CV-1 cells as described^{12,18} and modified for automation in 96-well plates⁷. All transfections were performed on a Beckman Biomek Automated Workstation. Transfections contained 5.0 ng of receptor expression plasmid vector (or control RS-CAT), 50 ng of the reporter luciferase plasmid, 50 ng of pRS β GAL (β -galactosidase) as an internal control, and 90 ng of carrier plasmid pGEM. Cells were transfected for 6 h, washed to remove DNA precipitates, and treated with hormones (1×10^{-6} M clofibric acid; 1×10^{-6} M 9-*cis* RA) for 36 h. Cell extracts were subsequently prepared and assayed for luciferase and β -galactosidase activities⁷. All experiments were done in triplicate in at least two independent experiments and were normalized for transfection efficiency by using β -galactosidase as the internal control.

response element (Fig. 3b, lanes 6–8). The observation that the PPAR-RXR α complex binds a sequence element found in the promoter of the 3-ketoacyl-CoA thiolase gene is particularly interesting, because this gene encodes another of the peroxisomal β -oxidation enzymes and is induced by peroxisomal proliferators¹⁹. The presence of DR-1-like HREs upstream of the genes for both peroxisomal enzymes suggests that the PPAR-RXR α complex is involved in their coregulation.

FIG. 2 Direct interactions between PPAR and RXR α in the absence or presence of DNA. *a*, PPAR and RXR α form a complex in solution. Immunoprecipitation was done using *in vitro*-synthesized [³⁵S]methionine-labelled PPAR in the presence of 150 ng of either bacterially expressed RXR α (lane 2) or control glutathione S-transferase (GST) (lane 1). Polyclonal antiserum prepared against RXR α was used. The position of immunoprecipitated PPAR is indicated by an arrow. Under identical conditions we failed to observe RXR α interactions with radiolabelled glucocorticoid receptor (data not shown, and ref. 14). *b*, PPAR and RXR α bind cooperatively to the PPRE. Gel mobility shift assays were done using *in vitro*-synthesized PPAR and/or RXR α as indicated, and ³²P-labelled PPRE oligonucleotide. Pre-immune (PI) or polyclonal antiserum prepared against RXR α (RXRab) were included in the reactions as indicated.

METHODS. PPAR and RXR α RNA was prepared and translated in rabbit reticulocyte lysates as directed by the supplier (Promega). Coimmunoprecipitation reactions (20 μ l) included 5 μ l of *in vitro*-synthesized [³⁵S]methionine-labelled PPAR and 150 ng of either bacterially expressed GST-RXR fusion protein¹² or GST alone in 20 mM Tris, pH 8.0. Proteins were incubated for 20 min on ice before the addition of 5 μ l polyclonal antisera prepared against an RXR α peptide (amino acids 214–229). Antigen-antibody complexes were collected by the addition of protein A-Sepharose (Pharmacia) and the immunocomplexes washed three times with 400 μ l RIPA buffer (10 mM Tris, pH 8.0, 150 mM NaCl, 1% Triton X-100, 1% sodium deoxycholate). Immunoprecipitated complexes were resolved by SDS-polyacrylamide gel electrophoresis on 10% gels which were then fixed in 30% methanol/10% acetic acid, dried, and subjected to autoradiography. Gel mobility shift assays (20 μ l) contained 10 mM Tris, pH 8.0, 40 mM KCl, 0.05% NP-40, 6% glycerol, 1 mM DTT, 0.2 μ g poly(dI-dC) and *in vitro*-synthesized PPAR (2.5 μ l) and RXR α (2.5 μ l) as indicated in the figure legends. The total amount of

The physiological significance of the PPAR-RXR α heterodimer is supported by the observations that both receptors are most highly expressed in the liver and kidney (data not shown, and refs 5 and 6), and that these tissues are the principal sites of action of peroxisome proliferators²⁰. Thus extracts prepared from these tissues should presumably contain factors that promote PPAR binding to the PPRE. As shown in Fig. 4, liver nuclear extracts have PPRE binding activity (lane 4), but



reticulocyte lysate was maintained constant in each reaction (5 μ l) through the addition of unprogrammed lysate. After a 10-min incubation on ice, 0.5–1 ng of ³²P-labelled oligonucleotide was added and the incubation continued for a further 10 min. DNA-protein complexes were resolved on a 4% polyacrylamide gel in 0.5 $\times$ TBE (1 $\times$ TBE is 90 mM Tris, 90 mM boric acid, 2 mM EDTA, pH 8.0). Gels were dried and autoradiographed at -70° . The PPRE oligonucleotide was 5'-AGCTGGACCAGGACAAAGGTCACGT-TCAGCT-3'.

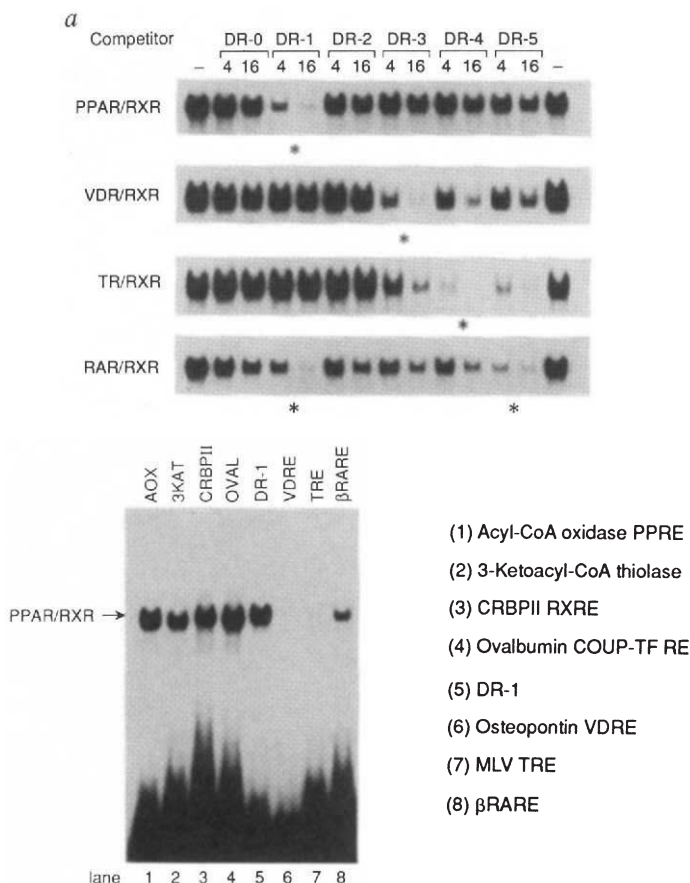


FIG. 3 Binding specificity of the PPAR-RXR α complex on synthetic and natural HREs. *a*, The binding specificities of the PPAR-RXR α , VDR-RXR α , TR-RXR α and RAR-RXR α complexes were tested by gel mobility shift analysis in the absence (–) or presence of a 4- or 16-fold molar excess of synthetic HREs, DR-0 to DR-5. Radiolabelled oligonucleotides encoding the PPRE, osteopontin VDRE, Moloney leukaemia virus (MLV) LTR TRE, and β -RARE were used as radiolabelled probes in binding assays involving the PPAR-RXR α , VDR-RXR α , TR-RXR α and RAR-RXR α complexes, respectively. Only the regions of the autoradiographs displaying the bound complexes are shown. Asterisks indicate reactions in which maximal levels of competition were observed. *b*, The binding specificity of the PPAR-RXR α complex was tested using radiolabelled oligonucleotides encoding the PPRE, 3-ketoacyl-CoA thiolase (3KAT), cellular retinol binding protein type II (CRBP II), chicken ovalbumin (OVAL), DR-1, osteopontin (VDRE), MLV LTR (TRE), and β -RARE (β-RARE) HREs. The sequences of these HREs are shown on the right.

METHODS. PPAR, RXR α , TR β , VDR and RAR α RNAs were prepared and

translated in rabbit reticulocyte lysates as directed by the supplier (Promega). Gel mobility shift assays were done as for Fig. 2. The 3-ketoacyl-CoA thiolase oligonucleotide was 5'-AGC-TCTCAGAGACCTTTGAACCA-CTTC-3'. The CRBP II-RXRE, β -RARE, MLV TRE, osteopontin VDRE, chicken ovalbumin HRE, and synthetic HRE direct repeat series oligonucleotides have been described^{12,13,18}.

AGGACA a AGGTCA

GGTTCA a AGGTCT

AGGTCA c AGGTCA c AGGTCA c AGTTCA

GTGTCA a AGGTCA

AGGTCA g AGGTCA

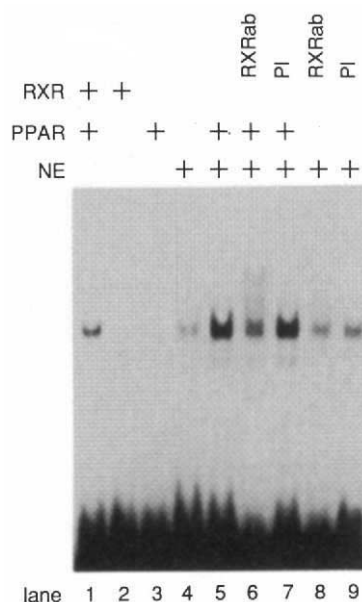
GGTTCA cga GGTTCa

GGGTCA tttc AGGTCC

GGTTCA ccgaa AGTTCA

FIG. 4 RXR α present in liver nuclear extracts enhances PPAR binding to the PPRE. Gel mobility shift assays were prepared using 32 P-labelled PPRE oligonucleotide and *in vitro*-synthesized PPAR and/or RXR α and 0.5 μ g of rat liver nuclear extract (NE) as indicated. Pre-immune (PI) or polyclonal antiserum prepared against RXR α (RXRab) was included in the reactions.

METHODS. Gel mobility shift assays were done as described for Fig. 2.



addition of *in vitro*-synthesized PPAR results in a marked enhancement in DNA binding (Fig. 4; compare lanes 4 and 5), demonstrating a cooperative activity in the extract. Addition of RXR α -specific antiserum to the reaction resulted in a reduction in complex formation to roughly the level in nuclear extracts alone, and the concomitant formation of a new complex with reduced electrophoretic mobility (Fig. 4, lane 6). These results demonstrate that RXR α is the main protein in liver responsible for enhancing PPAR binding to the PPRE *in vitro*.

In summary, our results demonstrate that peroxisome proliferators and 9-*cis* retinoic acid regulate an overlapping set of target genes through the PPAR-RXR α complex. The observation that both retinoids and peroxisome proliferators stimulate transcription of the genes encoding the peroxisomal β -oxidation enzymes^{3,4} suggests that 9-*cis* retinoic acid may function as a hypolipidaemic agent through activation of the PPAR-RXR α heterodimer. Finally, our demonstration that 9-*cis* retinoic acid is present in the liver and kidney of mouse⁷ is consistent with this molecule serving as an endogenous signal in the modulation of peroxisomal-mediated metabolism of fatty acids and other lipids. □

Received 6 April; accepted 8 July 1992.

1. Lazarow, P. B. & Fujiki, Y. A. *Rev. Cell Biol.* **1**, 489-530 (1985).
2. Vamecq, J. & Draye, J. P. *Essays Biochem.* **24**, 115-225 (1989).
3. Reddy, J. K. & Lalwani, N. D. *Crit. Rev. Toxicol.* **12**, 1-58 (1983).
4. Hertz, R. & Bar-Tana, J. *Biochem. J.* **281**, 41-43 (1992).
5. Isseman, I. & Green, S. *Nature* **347**, 645-650 (1990).
6. Mangelsdorf, D. J., Ong, E. S., Dyck, J. A. & Evans, R. M. *Nature* **345**, 224-229 (1990).
7. Heyman, R. A. *et al. Cell* **68**, 397-406 (1992).
8. Levin, A. A. *et al. Nature* **355**, 359-361 (1992).
9. Osumi, T., Wen, J.-K. & Hashimoto, T. *Biochem. biophys. Res. Commun.* **175**, 866-871 (1991).
10. Tugwood, J. D. *et al. EMBO J.* **11**, 433-439 (1992).
11. Dreyer, C. *et al. Cell* **68**, 879-887 (1992).
12. Mangelsdorf, D. J. *et al. Cell* **66**, 555-561 (1991).
13. Kliewer, S. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 1448-1452 (1992).
14. Kliewer, S. A., Umehara, K., Mangelsdorf, D. J. & Evans, R. M. *Nature* **355**, 446-449 (1992).
15. Yu, V. C. *et al. Cell* **67**, 1251-1266 (1991).
16. Leid, M. *et al. Cell* **68**, 377-395 (1992).
17. Zhang, X.-k. *et al. Nature* **355**, 441-446 (1992).
18. Umehara, K., Murakami, K. K., Thompson, C. C. & Evans, R. M. *Cell* **65**, 1255-1266 (1991).
19. Hijikata, M. *et al. J. Biol. Chem.* **265**, 4600-4606 (1990).
20. Nemali, M. R. *et al. Cancer Res.* **48**, 5316-5324 (1988).

ACKNOWLEDGEMENTS. We thank T. Perlmann for discussion; J. Dyck for antisera against RXR α ; D. Mangelsdorf for the RS-RXR α expression construct; T. Berger and members of New Leads Discovery for help with transfections assays; M. McKeown, H. Sucoy and D. Mangelsdorf for critically reading the manuscript, and E. Stevens for preparing it. S.A.K. is a Fellow of the Jane Coffin Childs Memorial Fund for Medical Research; K.U. is a Research Associate and R.M.E. is an Investigator of the Howard Hughes Medical Institute at the Salk Institute for Biological Studies. This work was supported by the National Institutes of Health and the Mathers Foundation.

Prediction of the structure of a receptor-protein complex using a binary docking method

Barry L. Stoddard & Daniel E. Koshland Jr

Division of Biochemistry, Department of Molecular and Cellular Biology, 534 Barker Hall, University of California, Berkeley, California 94720, USA

To validate procedures of rational drug design, it is important to develop computational methods that predict binding sites between a protein and a ligand molecule. Many small molecules have been tested using such programs, but examination of protein-protein and peptide-protein interactions has been sparse. We were able to test such applications once the structures of both the maltose-binding protein¹ (MBP) and the ligand-binding domain of the aspartate receptor², which binds MBP, became available. Here we predict the binding site of MBP to its receptor using a 'binary docking' technique in which two MBP octapeptide sequences containing mutations that eliminate maltose chemotaxis are independently docked to the receptor. The peptides in the docked solutions superimpose on their original positions in the structure of MBP and allow the formation of an MBP-receptor complex. The consistency of the computational and biological results supports this approach for predicting protein-protein and peptide-protein interactions.

The structure of the *Salmonella* receptor (which does not bind MBP) was used to create a model of the homologous receptor domain (which does bind MBP) from *Escherichia coli*. Using information from mutational studies, octapeptides from MBP were selected (Fig. 1) surrounding mutations that eliminate maltose chemotaxis^{3,5}. The peptides were then computationally annealed to their optimal positions on the aspartate receptor using the procedure described in ref. 6. The entire surface of the ligand-binding domain of the receptor was used as a target for the peptides. Comparison of the docked peptides with the MBP crystal structure provided an internal check (Fig. 2).

The backbone conformation of the peptides was kept the same as in the crystal structure and side chains were unrestrained.

MBP peptide 1:

Gln - Val - Ala - Ala - [Thr] - Gly - [Asp] - Gly - Pro
49 53 55

MBP peptide 2:

Trp - Tyr - Ala - Val - Arg - [Thr] - Ala - Val - Ile
340 345

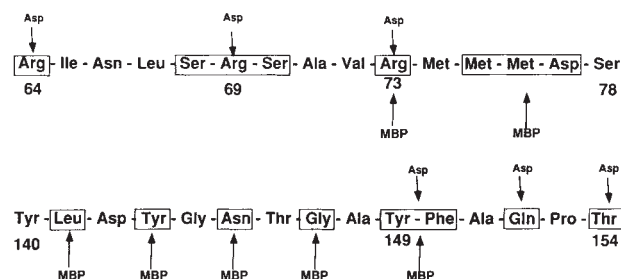


FIG. 1 Peptide sequences from maltose-binding protein (upper panel) and aspartate receptor (lower panel) implicated through mutational analysis as important for chemotaxis to maltose, and presumably for association of the binding protein and receptor. In the case of the aspartate receptor protein, substitutions shown above the actual sequence affect aspartate chemotaxis, and those below the sequence affect maltose chemotaxis. Note the adjacent but non-identical nature of these two sets of mutations.

We performed five experiments on each of the two peptides (residues 49–57 and 340–348). The results are shown in Fig. 3. Peptide 2 gave the same docked solution in three of five runs and a bound energy of $-15.2 \text{ kcal mol}^{-1}$. Peptide 1 minimized the same solution in four of five runs but its final energy was high ($2,411 \text{ kcal mol}^{-1}$), reflecting the fact that this peptide interacts with an area of the receptor that is partially disordered in

the X-ray structure². The distance between the docked peptides is $30 \text{ \AA}$. The distance between the same peptides in MBP (complexed to maltose) is $29 \text{ \AA}$. Superposition of the backbone atoms of residues 340–348 in maltose-binding protein with peptide 2 after docking leads to an overlap of peptide 1 with its corresponding residues in MBP (r.m.s. $2 \text{ \AA}$). Restrained minimization of the resulting complex structure produces a final solution of

FIG. 2 The rationale and methodology of the binary docking strategy. Two proteins (P1 and P2) that have been solved crystallographically are known to associate, and the only clue to the possible sites of binding are amino-acid mutations at locations X and Y which prevent the biological response. Peptides from the regions around X and Y (XP and YP) are exposed to the entire protein P2 and independently docked. This identifies a pair of binding sites (X' and Y'). The distance apart and relative alignment of the two docked peptides can then be compared to the same sequences in the crystal structure of protein 1. If the geometry between docked peptides does not agree with the crystal structure of protein P1, or the predicted complex structure causes serious physical overlap and improper contacts, the evidence indicates the calculations are not correct. Automated docking of substrates to proteins³ combines simulated annealing with rapid energy evaluation¹⁰. Partial charges are assigned to each atom in the receptor structure and in the probe peptides using the Gasteiger method¹¹ in the molecular modelling package QUANTA. Grids of molecular affinity potentials¹² are calculated which encase the protein surface for each type of atom in the substrate molecules; the grids were calculated in steps of $0.5 \text{ \AA}$ in each direction for $30 \text{ \AA}$ on each side (producing a $2,700 \text{ \AA}^3$ box). The molecules are translated into the grid, with an initial energy for the system of $100 \text{ kcal mol}^{-1}$. Two hundred cycles of slow cooling were performed with a maximum of 30,000 accepted or rejected steps per cycle, as determined previously³. The maximum translation for the peptides was $0.4 \text{ \AA}$ and the maximum bond rotation was 5° per step.

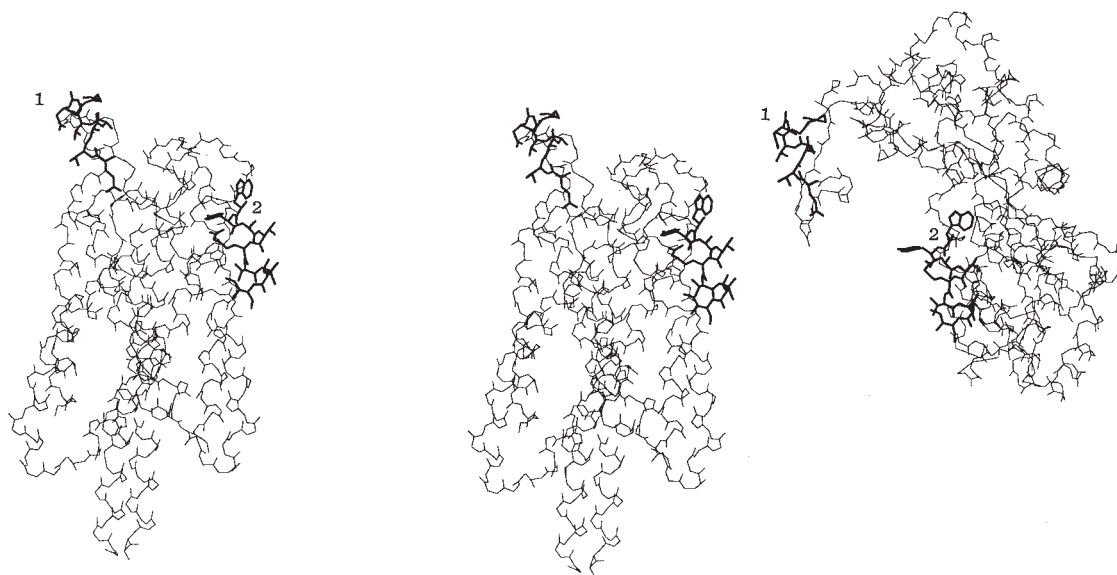
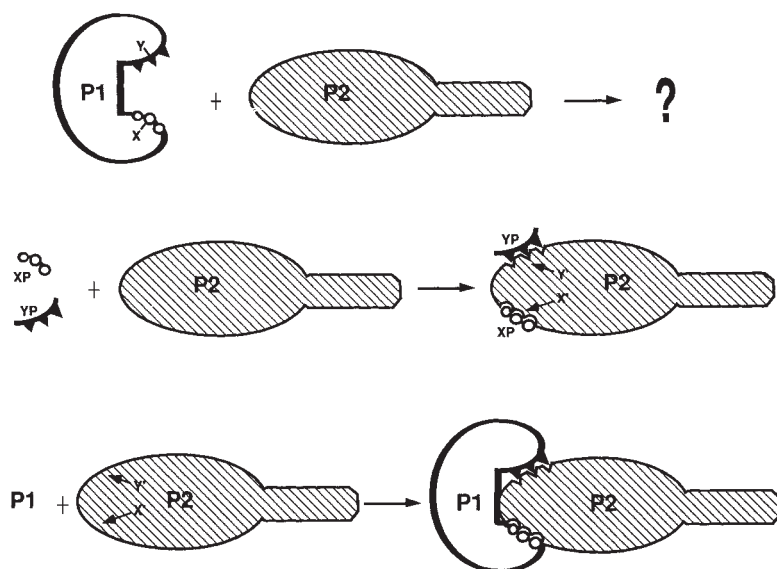
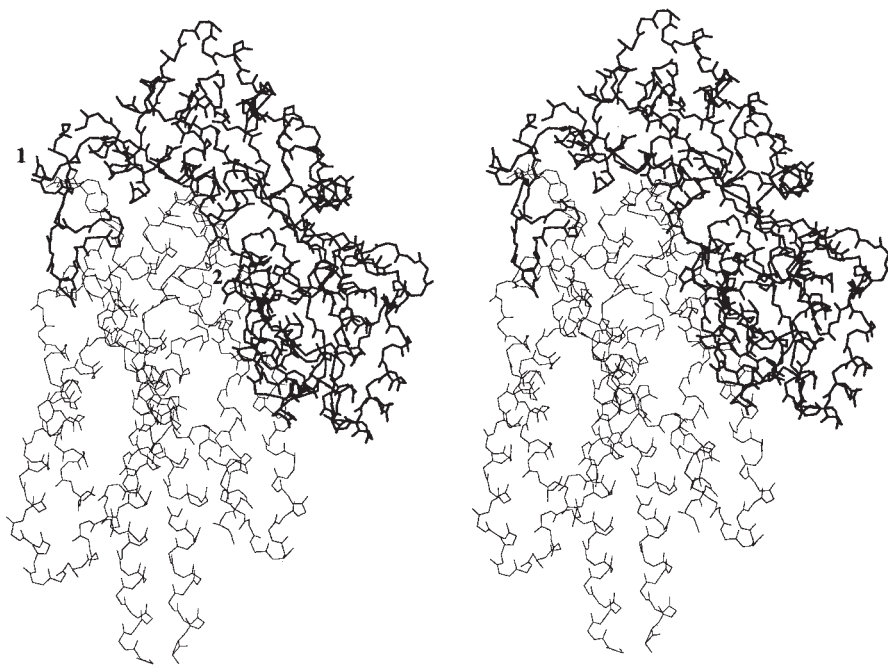


FIG. 3 Stereo view of the two peptides from MBP (heavy bonds) in their computationally docked positions on the aspartate receptor (light bonds). MBP (from the computationally derived receptor complex) with the same peptides highlighted is shown for comparison. Peptide 2 packs against the helices 2 and 4' at the receptor dimer interface, in contact with residues 73–85 and 148'–152' (previously implicated as involved in MBP interactions) of aspartate receptor protein. Peptide 1 is in contact with residues 73' to 81' of the receptor, interacting with a flexible loop of the receptor which has also been shown to bind MBP. Residues 53 and 55 from MBP have formed an interaction with residues 75', 76' (methionine), and 149 (tyrosine) of the receptor. Superposition of the backbone atoms of residues 340–348 in MBP with peptide 2 after docking leads to an almost direct overlap of

peptide 1 with its corresponding residues in MBP (r.m.s. $2 \text{ \AA}$). After minimization of the complex, the loop in the N-terminal domain of MBP containing peptide 1 has swung out slightly from the binding protein, yielding its structure in the final complex. Control docking experiments with polyaniline peptides in the same backbone conformations do not identify sites as in Fig. 2. Substitution for amino acids that participate in MBP binding as shown in Fig. 1 reduce the final binding energy for peptide 2 by about 10 times (but give the same result), and altogether eliminate docking of peptide 1. Interestingly, the original peptides will dock computationally to *Salmonella* aspartate receptor (which does not bind MBP); the receptor sequences indicate that prevention of maltose taxis in *Salmonella* is due to other unidentified residues.

FIG. 4 Stereo diagram of the complex of the maltose-binding protein (heavy bonds) bound to the dimeric structure of the periplasmic domain of the aspartate receptor (light bonds). Only the protein backbone is shown for clarity. The MBP monomer is bound at the receptor-dimer interface. In the initial model of the complex produced by substitution of intact maltose-binding protein for the docked peptides, minor steric conflict was observed between the two proteins in only two locations involving surface loops of MBP and aspartate receptor protein. Restrained minimization of the resulting complex structure produces a final satisfactory solution for geometry and energy. Among the previous genetic and biochemical observations regarding maltose taxis as mediated by the aspartate receptor protein, those that support the predicted model include the participation of Arg 73 and Tyr 149 from the receptor in interactions with both aspartate and MBP, Met 75 and 76 in MBP binding, and the ability of the receptor to support independent and additive responses to aspartate and maltose.



good geometry (Fig. 4). On the basis of the agreement of two independent solutions with the structure of MBP, and the complementarity of the resulting complex structure, we conclude that our protocol has yielded a correct solution.

The predicted structure of the complex contains details that agree with observations concerning maltose chemotaxis. One result of genetic experiments shows that, of the five residues that bind aspartate, two (Arg 73 and Tyr 149) are also important for response to maltose⁵. Mutants of Arg 73 suppress the effect of the maltose taxis defects associated with specific mutations in MBP (ref. 4). In the model, Arg 73 is the only aspartate-binding residue that is directly involved in MBP binding, through hydrogen bonds formed between its side chain and the backbone of the binding protein. The side chains of Arg 64, 69, Tyr 149, and Thr 154 do not participate in MBP binding. This observation is true for both aspartate-binding sites. Tyrosine 149 is involved in protein binding through interactions that do not involve the aromatic ring.

Three residues from MBP have been identified as having special importance in chemotaxis³⁻⁵: Thr 53 and Asp 55 (peptide 1) and Thr 345 (peptide 2). In our model, Thr 345 is found in a helix that packs into the dimer interface between helices 2 and 4' of the receptor. Although in the docking runs the ultimate orientation of this peptide is unbiased, the docked solution shows Thr 345 pointing in towards the aspartate receptor protein

structure and interacting with helix 2 of the receptor. The other two sites on MBP (53 and 55) interact with residues Tyr 149, Phe 150, and Met 75 and 76 of the receptor. A large number of specific structural interactions predicted in previous mutational studies are thus confirmed by this docked model.

The receptor can respond simultaneously and independently to both aspartate and maltose⁷. This necessitates separate binding of both aspartate and maltose-binding protein and additive signalling events through conformational changes in the receptor structure^{7,8}. In the receptor complex, one of the two available aspartate-binding sites is buried under the MBP-receptor site, whereas the second is accessible. Separate evidence supports aspartate binding to half of the sites in the absence of MBP. This implies that aspartate is only capable of binding a single accessible site before or after the binding of MBP, which would explain the independent additive effect⁷. In the case of a mechanism involving conformational changes within individual monomers, signalling and adaptation could be induced within one receptor subunit by a single bound aspartate⁹ and a second signal transduced through the second subunit in response to the binding of MBP. It is possible that the receptor has evolved to prevent saturation of its signalling potential by high aspartate concentrations, which would negate the ability to display a separate signal in response to a second stimulating ligand such as MBP. □

Received 13 May; accepted 8 July 1992.

1. Spurlino, J. C., Lu, G.-Y. & Quijcho, F. A. *J. Biol. Chem.* **266** (8), 5202-5219 (1991).
2. Milburn, M. V. *et al. Science* **254**, 1342-1347 (1991).
3. Manson, M. D. & Kossman, M. *J. Bact.* **165**, 34-40 (1986).
4. Kossman, M., Wolff, C. & Manson, M. D. *J. Bact.* **170**, 4516-4521 (1988).
5. Gardina, P., Conway, C., Kossman, M. & Manson, M. D. *J. Bact.* **174**, 1528-1526 (1992).
6. Goodsell, D. & Olson, A. *Proteins Struct. Funct. Genet.* **8**, 195-202 (1991).
7. Mowbray, S. L. & Koshland, D. E. *Jr Cell* **50**, 171-180 (1987).

8. Falke, J. J. & Koshland, D. E. *Jr Science* **237**, 1596-1600 (1987).
9. Milligan, D. & Koshland, D. E. *Jr Science* **254**, 1651-1654 (1991).
10. Kirkpatrick, S., Gelatt, C. D. Jr & Vecchi, M. P. *Science* **220**, 671-680 (1983).
11. Gasteiger, J. & Marsili, M. *Tetrahedron Lett.* **34**, 3181-3184 (1978).
12. Goodford, P. J. *J. med. Chem.* **28**, 849-857 (1985).

ACKNOWLEDGEMENTS. We thank the laboratory of F. Quijcho for MBP coordinates, M. Manson for preprints of genetic work on MBP, and D. Goodsell and A. Olson for FORTRAN code for the docking protocol and for technical advice. This work was supported by the NIH; B.L.S. is a fellow of the Helen Hay Whitney Biomedical Research Foundation.